

STIC-Biotech/ChemLib

79425

09/782,874

From: Chan, Christina  
Sent: Tuesday, November 05, 2002 9:32 AM  
To: Helmer, Georgia; STIC-Biotech/ChemLib  
Subject: RE: Please Rush Sequence Search-09/782,874--

Point of Contact:  
Mona Smith  
Technical Information Specialist  
CM1 6A01  
Tel: 308-3278

Please rush. Thanks Chris

Chris Chan  
TC 1600 New Hire Training Coordinator and SPE 1644  
308-3973  
CM-1, 9B19

CRFE

-----Original Message-----

From: Helmer, Georgia  
Sent: Monday, November 04, 2002 6:24 PM  
To: Chan, Christina  
Subject: Please Rush Sequence Search-09/782,874--

Could you please do a sequence search and an oligo search on SEQ ID No 1 and 2?

On the results for the oligo searches, could you please print out more of the results? Like up to the first 50?

This is for the inhouse and commerce data bases. (no interference search though)

I need this case for this biweek.

Thanks in advance, Georgia L. Helmer Ph.D.  
Patent Examiner  
Crystal Mall 1, 9D14  
AU 1638  
703-308-7023  
Georgia.Helmer@USPTO.gov

16 coning bases  
= 0.42%  
0.42%

Seq ID 1 = 37.31 nuc

Searcher: M. Smith  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 11/04/02  
Date Completed: 11/06/02  
Searcher Prep/Review: 10  
Clerical: \_\_\_\_\_  
Online time: 15

TYPE OF SEARCH:  
NA Sequences: 2  
AA Sequences: 2  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 19:16:42 ; Search time 142 seconds  
(without alignments)  
8057.824 Million cell updates/sec

Title: US-09-782-874-1  
Perfect score: 3731  
Sequence: 1 GAATATCTTCTTACTTACTT.....AGTTTCATCTTCTTCTTCTTAA 3731

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/6C\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/6D\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3731	100.0	3731	4	US-08-811-583-1
2	57.2	1.5	906	1	Sequence 1, Appl
3	53.4	1.4	7218	1	Sequence 14, Appl
4	40	1.1	162450	4	Sequence 1, Appl
5	36.8	1.0	2277	1	Sequence 2, Appl
6	36.8	1.0	2277	1	Sequence 2, Appl
7	36.8	1.0	2277	2	Sequence 2, Appl
8	36.6	1.0	7218	1	Sequence 14, Appl
9	36.4	1.0	3191	4	Sequence 194, Appl
10	35.4	0.9	17341	4	Sequence 1, Appl
11	35	0.9	1248	4	Sequence 342, Appl
12	35	0.9	1275	2	Sequence 4, Appl
13	35	0.9	1275	4	Sequence 4, Appl
14	35	0.9	1458	4	Sequence 19, Appl
15	35	0.9	1458	4	Sequence 19, Appl
16	35	0.9	1796	2	Sequence 17, Appl
17	35	0.9	1796	4	Sequence 17, Appl
18	35	0.9	1845	2	Sequence 5, Appl
19	35	0.9	1845	4	Sequence 5, Appl
20	35	0.9	2046	2	Sequence 6, Appl
21	35	0.9	2046	4	Sequence 6, Appl
22	35	0.9	5613	2	Sequence 1, Appl
23	35	0.9	111282	4	Sequence 3, Appl
24	34.8	0.9	1028	4	Sequence 222, Appl
25	34.8	0.9	6002	4	Sequence 4, Appl
26	34.8	0.9	19446	4	Sequence 51, Appl
27	34.6	0.9	1883	4	Sequence 3, Appl

28	34.6	0.9	19390	4	US-08-961-527-86	Sequence 86, Appl
29	34.4	0.9	738	4	US-09-004-838-62	Sequence 62, Appl
30	34.2	0.9	1003	4	US-09-221-017B-395	Sequence 395, Appl
31	34.2	0.9	2694	1	US-08-465-995A-1	Sequence 1, Appl
32	34.2	0.9	2694	2	US-08-465-994C-1	Sequence 1, Appl
33	34.2	0.9	2694	2	US-08-966-145-1	Sequence 1, Appl
34	34.2	0.9	2760	1	US-08-101-593-1	Sequence 1, Appl
35	33.8	0.9	258	4	US-09-345-882-21	Sequence 21, Appl
36	33.8	0.9	1867	4	US-09-377-557-3	Sequence 3, Appl
37	33.8	0.9	2889	4	US-09-134-001C-1660	Sequence 1660, Appl
38	33.6	0.9	2870	4	US-09-134-078-59	Sequence 59, Appl
39	33.6	0.9	3707	1	US-08-118-101A-1	Sequence 1, Appl
40	33.6	0.9	10223	4	US-08-961-527-73	Sequence 73, Appl
41	33.4	0.9	1939	1	US-08-212-188-3	Sequence 3, Appl
42	33.4	0.9	1939	3	US-08-970-725-3	Sequence 3, Appl
43	33.4	0.9	1939	5	PCT-US95-02708-3	Sequence 3, Appl
44	33.4	0.9	2026	3	US-08-755-587-26	Sequence 26, Appl
45	33.2	0.9	3636	4	US-09-134-001C-1983	Sequence 1983, Appl

## ALIGNMENTS

## RESULT 1

US-08-811-583-1  
; Sequence 1, Application US/08811583  
; Patent No. 6218142  
; GENERAL INFORMATION:  
; APPLICANT: Wassenecker, Michael  
; APPLICANT: Riedel, Leonhard  
; APPLICANT: Schiebel, Winfried  
; APPLICANT: Sanger, Heinz  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMIC ACTIVITY OF AN  
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RDRP)  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & NEAVE  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/811,583  
; FILING DATE: 05-MAR-1997  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley, James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: MPG-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3731 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ORIGINAL SOURCE:  
; ORGANISM: Tomato  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 194...3535  
US-08-811-583-1



Db 2101 TGATCTTTTGAAGGCGCAGGAGGCTTTGGAATTGATGCTCTCTGGAGAGACACTAATAT 2160  
QY 2161 TCTCAAGGCAATGCTAACTGTGTTTAAAGCCTGATGCTGAGCCCTTTCTTTCAATGAT 2220  
Db 2161 TCTCAAGGCAATGCTAACTGTGTTTAAAGCCTGATGCTGAGCCCTTTCTTTCAATGAT 2220  
QY 2221 GTTTCGAACCTTCCGCGCATCCAGTGTGCTGATTTGCGGACTAGATCAAGAAATATTTAT 2280  
Db 2221 GTTTCGAACCTTCCGCGCATCCAGTGTGCTGATTTGCGGACTAGATCAAGAAATATTTAT 2280  
QY 2281 TCCAAATGGAAGAACATGATGGAATGTTGGATGAATCCAGAACCTTGGATATGATGCTCA 2340  
Db 2281 TCCAAATGGAAGAACATGATGGAATGTTGGATGAATCCAGAACCTTGGATATGATGCTCA 2340  
QY 2341 GGTGTTTCTTCAAGTTTACTGTGCTGACATGGAGAGTCTTTCGACGATTTTACATCCATT 2400  
Db 2341 GGTGTTTCTTCAAGTTTACTGTGCTGACATGGAGAGTCTTTCGACGATTTTACATCCATT 2400  
QY 2401 TAATAACAGCAGATCCCAACACAGTAATTCATCTGAAGGGAATGTGGTGTGTTCCAAA 2460  
Db 2401 TAATAACAGCAGATCCCAACACAGTAATTCATCTGAAGGGAATGTGGTGTGTTCCAAA 2460  
QY 2461 AAATCCATGCTTGCATCTGCTGATATTCGTTGTTTAAAGGCTGTAATGTTTCGAGCGCT 2520  
Db 2461 AAATCCATGCTTGCATCTGCTGATATTCGTTGTTTAAAGGCTGTAATGTTTCGAGCGCT 2520  
QY 2521 GCACACAGTGTGATGTTGTTTATTCCTCAGAAAGAAAGACCTCATCCGAATGA 2580  
Db 2521 GCACACAGTGTGATGTTGTTTATTCCTCAGAAAGAAAGACCTCATCCGAATGA 2580  
QY 2581 ATGTTCTGGAGTGTGATGTTGGATATCTACTTTGTTTGGGATCAACACATGAT 2640  
Db 2581 ATGTTCTGGAGTGTGATGTTGGATATCTACTTTGTTTGGGATCAACACATGAT 2640  
QY 2641 CCCGCAAGGCAAGTCCAGCGATGGAATATCCCTCAGACCCAGCAGATATGATGATGATG 2700  
Db 2641 CCCGCAAGGCAAGTCCAGCGATGGAATATCCCTCAGACCCAGCAGATATGATGATGATG 2700  
QY 2701 TGATCTCAATTTGAGGAGTGTGAGTACTTCCACCACTATATTTGTAATCAGAGTTT 2760  
Db 2701 TGATCTCAATTTGAGGAGTGTGAGTACTTCCACCACTATATTTGTAATCAGAGTTT 2760  
QY 2761 GGAATATAGCAATGCCATGCTGTTTATTCGACAGAGAACCTGATATGCCATGAG 2820  
Db 2761 GGAATATAGCAATGCCATGCTGTTTATTCGACAGAGAACCTGATATGCCATGAG 2820  
QY 2821 TGATCCATGCAAAACCTGCTGAGCTTTTCAATTCGAGTGGACTTTCCAAAGACTGG 2880  
Db 2821 TGATCCATGCAAAACCTGCTGAGCTTTTCAATTCGAGTGGACTTTCCAAAGACTGG 2880  
QY 2881 TGTTCCTCCGCTGAAATACCATCTCAGTTGCGCCCTTAAAGATACCCAGCTTCATGATAA 2940  
Db 2881 TGTTCCTCCGCTGAAATACCATCTCAGTTGCGCCCTTAAAGATACCCAGCTTCATGATAA 2940  
QY 2941 GCGGACAGACCCAGCTATATCTCAGAAAGAGTTATTGGAAGCTTTTCAGAAAGTGAA 3000  
Db 2941 GCGGACAGACCCAGCTATATCTCAGAAAGAGTTATTGGAAGCTTTTCAGAAAGTGAA 3000  
QY 3001 GGACAAAGCAGCTCAGGCTAGCTATCGGACCTTCCAGAGAGTGTTCGAAGGATC 3060  
Db 3001 GGACAAAGCAGCTCAGGCTAGCTATCGGACCTTCCAGAGAGTGTTCGAAGGATC 3060  
QY 3061 ATATGATGCTGATGGAAGTGTGATTTGAGATTTACATTCAGCAAGCTTTTGACAT 3120  
Db 3061 ATATGATGCTGATGGAAGTGTGATTTGAGATTTACATTCAGCAAGCTTTTGACAT 3120  
QY 3121 CAAAACTGAATATGACAAACAGCTGGGTAAATTTAATGAGTACTATGGCATAAAACAGA 3180  
Db 3121 CAAAACTGAATATGACAAACAGCTGGGTAAATTTAATGAGTACTATGGCATAAAACAGA 3180  
QY 3181 GCGTGAATATCTAGTGTGCAATATGAGGAGTCAAAACCTTTGACCGGAGAAAGA 3240  
Db 3181 GCGTGAATATCTAGTGTGCAATATGAGGAGTCAAAACCTTTGACCGGAGAAAGA 3240

QY 3241 TGCTGAGCCCATTAGTGTGCTGTGAGGCGCTTGAGAAAGGAGGCAAGAGCTGGTTCAA 3300  
Db 3241 TGCTGAGCCCATTAGTGTGCTGTGAGGCGCTTGAGAAAGGAGGCAAGAGCTGGTTCAA 3300  
QY 3301 GAGCGGTAATGATATAGATGACATGTTACAAAGGCTTGGCTTGTGATCCAGCTTACATA 3360  
Db 3301 GAGCGGTAATGATATAGATGACATGTTACAAAGGCTTGGCTTGTGATCCAGCTTACATA 3360  
QY 3361 TCATCTACATATTTGGGTTGCTACATCAGGCTTGAAGAGCTCATTTCAATAGCTT 3420  
Db 3361 TCATCTACATATTTGGGTTGCTACATCAGGCTTGAAGAGCTCATTTCAATAGCTT 3420  
QY 3421 TCCCTGTGTTTATGACCAAGCTTAATCCAGATTAAGAGGCAAAAGCAGCTTAACAGGCC 3480  
Db 3421 TCCCTGTGTTTATGACCAAGCTTAATCCAGATTAAGAGGCAAAAGCAGCTTAACAGGCC 3480  
QY 3481 AGTTCTCAACTGTGATCTCTCAGGCTCAACTGAGTCAGATTAAGTGTGAATGAGA 3540  
Db 3481 AGTTCTCAACTGTGATCTCTCAGGCTCAACTGAGTCAGATTAAGTGTGAATGAGA 3540  
QY 3541 TTCCAGTCGAGCGTAAAGCTGATATATATTAATAGGCTGTGATCAATAGAAAC 3600  
Db 3541 TTCCAGTCGAGCGTAAAGCTGATATATATTAATAGGCTGTGATCAATAGAAAC 3600  
QY 3601 TGTATGATGTTGATGATCTTTTGTCTTAAACTGCAATGAGCTGCAACATATGTC 3660  
Db 3601 TGTATGATGTTGATGATCTTTTGTCTTAAACTGCAATGAGCTGCAACATATATGTC 3660  
QY 3661 AGTACTCTAAGAAACAGATGTACAGCTAAGTACTATATGATGTTGATGTTGATC 3720  
Db 3661 AGTACTCTAAGAAACAGATGTACAGCTAAGTACTATATGATGTTGATGTTGATC 3720  
QY 3721 TTTCTTCTAAA 3731  
Db 3721 TTTCTTCTAAA 3731

RESULT 2

US-08-100-874-1/c  
; Sequence 1, Application US/08100874  
; Patent No. 5498533  
; GENERAL INFORMATION:  
; APPLICANT: Poovaiiah, B. W.  
; APPLICANT: Takezawa, D.  
; APPLICANT: Han, T. J.  
; APPLICANT: An, G. H.  
; TITLE OF INVENTION: Control of Growth and Development of  
; TITLE OF INVENTION: Potato Plants  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harness, Dickey & Pierce  
; STREET: P.O. Box 828  
; CITY: Bloomfield Hills  
; STATE: MI  
; COUNTRY: USA  
; ZIP: 48303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/08100, 874  
; APPLICATION NUMBER: US/08100, 874  
; FILING DATE: July 30, 1993  
; CLASSIFICATION: /800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, Deann F.  
; REFERENCE/DOCKET NUMBER: 7555-00004  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (313) 641-1600  
; TELEFAX: (313) 641-0270  
; TELEX: 287637 Harness UR



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:33:33 ; Search time 142 Seconds  
(without alignments)  
1616.454 Million cell updates/sec

Title: US-09-782-874-2  
Perfect score: 1114  
Sequence: 1 MGKTIQVGFPPYLLSAEVRK.....RPVNLSSLRQLSHRLVLRK 1114

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database :

- SPTREMBL\_21.\*
- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1114	100.0	1114	10	Q9ZR58
2	42	3.8	1116	10	Q9ZRY7
3	38	3.4	775	10	Q9ZRX3
4	23	2.1	513	10	Q9ZRX5
5	23	2.1	1107	10	Q9LQV2
6	18	1.6	107	10	Q9ZRX4
7	14	1.3	1133	10	Q82504
8	12	1.1	1196	10	Q9SG02
9	12	1.1	1196	10	Q9LKP0
10	10	0.9	1377	5	Q95WU3
11	9	0.8	583	17	Q26600
12	9	0.8	1069	5	Q9XYP1
13	9	0.8	1638	3	Q9C162
14	8	0.7	198	12	Q39857
15	8	0.7	265	3	Q12044
16	8	0.7	286	16	Q92Y23

17	8	0.7	352	16	Q97ES6
18	8	0.7	444	5	Q95ZS5
19	8	0.7	490	11	Q9CU85
20	8	0.7	515	10	Q9SP43
21	8	0.7	583	13	Q9DG87
22	8	0.7	583	13	Q9IAT4
23	8	0.7	598	4	Q9Y232
24	8	0.7	635	17	Q9EYH3
25	8	0.7	830	10	Q65482
26	8	0.7	864	5	Q9G064
27	8	0.7	1101	16	Q53347
28	8	0.7	1226	4	Q96S73
29	8	0.7	1283	4	Q96S72
30	8	0.7	1283	4	Q96S71
31	8	0.7	1344	10	Q8S921
32	8	0.7	1564	5	Q9XW49
33	8	0.7	1883	5	Q95ZG7
34	8	0.7	1964	5	Q95ZG6
35	8	0.7	3229	2	Q9ETK0
36	7	0.6	29	16	Q8Z6B7
37	7	0.6	36	6	Q9GKJ9
38	7	0.6	43	4	Q9HB48
39	7	0.6	66	16	Q8YNP2
40	7	0.6	70	5	Q9GUX8
41	7	0.6	72	5	Q9NAV2
42	7	0.6	72	5	Q9NAU5
43	7	0.6	72	5	Q9NAU4
44	7	0.6	72	5	Q9NAU3
45	7	0.6	72	5	Q9NAU2
46	7	0.6	72	5	Q9NAT7
47	7	0.6	78	6	Q9RTV1
48	7	0.6	79	16	Q8Z1J9
49	7	0.6	89	4	Q9NT43
50	7	0.6	104	16	Q98111

ALIGNMENTS

RESULT 1

Q9ZR58	PRELIMINARY;	PRT;	1114	AA.
AC	Q9ZR58:			
DT	01-MAY-1999 (TRENBLrel. 10, Created)			
DT	01-MAY-1999 (TRENBLrel. 10, Last sequence update)			
DT	01-JUN-2002 (TRENBLrel. 21, Last annotation update)			
DE	RNA-directed RNA polymerase (EC 2.7.7.48).			
GN	RDRP.			
OS	Lycopodium esculentum (Tomato).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
OC	Asteridae; euasterids I; Solanales; Solanaceae; Solanum.			
OX	NCBI_TaxID=4081;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=RUTGERS;			
RX	MEDLINE=99055198; PubMed=9836747;			
RA	Schibel W., Pellissier T., Riedel L., Thalmair S., Schibel R.,			
RA	Kempe D., Lottspeich F., Sanger H.L., Wassenegger M.;			
RT	"Isolation of an RNA-directed RNA polymerase-specific cDNA clone from			
RT	tomato."			
RL	Plant Cell 10:2087-2102(1998).			
DR	EMBL: Y10403; CAA71421.1; -			
DR	InterPro: IPR000504; RNA_rec_mot.			
DR	InterPro: IPR003955; RRM_2.			
DR	Pfam: PF00076; rrm; 1.			
DR	SMART: SM00362; RRM_2; 1.			
KW	Nucleotidyltransferase; RNA-directed RNA polymerase; Transferase.			
SQ	SEQUENCE 1114 AA; 126807 MW; B6D30ACD41DB37FF CRC64;			

Query Match 100.0%; Score 1114; DB 10; Length 1114;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKTIQVGFPPYLLSAEVAWSFLEKTYGTGTCVALEVKQSKGGSRAFAKQVQFADNISADK 60  
DB 1 MGKTIQVGFPPYLLSAEVAWSFLEKTYGTGTCVALEVKQSKGGSRAFAKQVQFADNISADK 60  
QY 61 IITLANNRLYFGSSYLKAWEMKTDIVOLRAYVDMQDITLNFQCOISDDKFAVLGSTEVS 120  
DB 61 IITLANNRLYFGSSYLKAWEMKTDIVOLRAYVDMQDITLNFQCOISDDKFAVLGSTEVS 120  
QY 121 IQFGIGLKKFFFFLLSSGSAKYKQLQSYENWQVVLHRYPGQNAQFLLIQLFGAPRIYKRL 180  
DB 121 IQFGIGLKKFFFFLLSSGSAKYKQLQSYENWQVVLHRYPGQNAQFLLIQLFGAPRIYKRL 180  
QY 181 ENSCYSFKEKTPDDQWTRTDFPPSWITGLSSSLCLOFRGRVRLPNFEESFFHYAERENNI 240  
DB 181 ENSCYSFKEKTPDDQWTRTDFPPSWITGLSSSLCLOFRGRVRLPNFEESFFHYAERENNI 240  
QY 241 TLOGFFFFYSOKSALYPNVQPEGISIPYKILFKISSLVQHCICPGPALNVYFFRLVDP 300  
DB 241 TLOGFFFFYSOKSALYPNVQPEGISIPYKILFKISSLVQHCICPGPALNVYFFRLVDP 300  
QY 301 RRRNVACIEHALEKLYIKKCCYDPVWLTQYDGYLKGROPKSPSITLDDGLVYVRRV 360  
DB 301 RRRNVACIEHALEKLYIKKCCYDPVWLTQYDGYLKGROPKSPSITLDDGLVYVRRV 360  
QY 361 LVTPCKYVFCGPEVNVSRVLRNYSIEDINFLRVSFVDEWEKLYSTDLLPKASTGSGVR 420  
DB 361 LVTPCKYVFCGPEVNVSRVLRNYSIEDINFLRVSFVDEWEKLYSTDLLPKASTGSGVR 420  
QY 421 TNYIERILSTLRKGFVIGDKKFEFLAFSSQLRDNYSVMFASRPLGTANDIRAWMGDFSQ 480  
DB 421 TNYIERILSTLRKGFVIGDKKFEFLAFSSQLRDNYSVMFASRPLGTANDIRAWMGDFSQ 480  
QY 481 IKNYAKAARLGQSGFSRSTLSVLRHEIEVDPVKVHGTSYVFSOGIGISGDFAHRYA 540  
DB 481 IKNYAKAARLGQSGFSRSTLSVLRHEIEVDPVKVHGTSYVFSOGIGISGDFAHRYA 540  
QY 541 SKCGLQTPSAFAQRYGKYGVGVDPDSSMKLSLRKSMKSYESDNKLDVLGWSKYQPC 600  
DB 541 SKCGLQTPSAFAQRYGKYGVGVDPDSSMKLSLRKSMKSYESDNKLDVLGWSKYQPC 600  
QY 601 YLNQLTLTLSTLGVKQDEVLEQKQEAVDQDLDAILHDSLKAQEALEMLSGENTNLIKAM 660  
DB 601 YLNQLTLTLSTLGVKQDEVLEQKQEAVDQDLDAILHDSLKAQEALEMLSGENTNLIKAM 660  
QY 661 LNCYKPDAPFELSMLOTPFRASKLDDLRSTRSIFINGRMTMGCLDESRTLEYGVFVQ 720  
DB 661 LNCYKPDAPFELSMLOTPFRASKLDDLRSTRSIFINGRMTMGCLDESRTLEYGVFVQ 720  
QY 721 FTGAGHGEFSDLLHPFNNSRSTNSFILKGNVYVAKNPCLHPGDIRVLKAVNVRAALHHV 780  
DB 721 FTGAGHGEFSDLLHPFNNSRSTNSFILKGNVYVAKNPCLHPGDIRVLKAVNVRAALHHV 780  
QY 781 DCVVFPQKGRPHNEGSGDLDGDIYFVCWQDMIPPROVQMEYPPAPSIQLDHDTV 840  
DB 781 DCVVFPQKGRPHNEGSGDLDGDIYFVCWQDMIPPROVQMEYPPAPSIQLDHDTV 840  
QY 841 EEVEEYFTNYIVNDSLGITIANHVVFADREPDMAMSDPCKKLAEFLSIAYDFPKTGPAE 900  
DB 841 EEVEEYFTNYIVNDSLGITIANHVVFADREPDMAMSDPCKKLAEFLSIAYDFPKTGPAE 900  
QY 901 IPSQLRPKEYDFMDKPKDTSYISERVIGKLFVRKVKQKAPQASSIAFTROVAPRSYDAD 960  
DB 901 IPSQLRPKEYDFMDKPKDTSYISERVIGKLFVRKVKQKAPQASSIAFTROVAPRSYDAD 960  
QY 961 MEVDGFEDYIDEAFDYKTEYDNKLGNLMDYGIKTEAEILSGGIMKASKTFDRRKDAEAI 1020  
DB 961 MEVDGFEDYIDEAFDYKTEYDNKLGNLMDYGIKTEAEILSGGIMKASKTFDRRKDAEAI 1020  
QY 1021 SVAVRALRKEARAFKRRNDIDMLPKASAWHYHTYHTYWGVCYNQGLKRAHIFISFPWCY 1080  
DB 1021 SVAVRALRKEARAFKRRNDIDMLPKASAWHYHTYHTYWGVCYNQGLKRAHIFISFPWCY 1080

QY 1081 YDOLIOIKKDKARNRPVNLSSLAQLSHRLVLK 1114  
DB 1081 YDOLIOIKKDKARNRPVNLSSLAQLSHRLVLK 1114  
RESULT 2  
Q9ZRY7  
ID Q9ZRY7 PRELIMINARY; PRT: 1116 AA.  
AC Q9ZRY7;  
DT 01-MAY-1999 (TREMELrel. 10, Created)  
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)  
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)  
DE RNA-directed RNA polymerase.  
GN RDRP.  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. PETIT HAVANA SRI; TISSUE=LEAF;  
RA Schiebel W., Pellissier T., Riedel L., Thalmeir S., Schiebel R.,  
RA Kempe D., Lottspeich F., Saenger H.L., Wassenegeger M.;  
RT "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from  
RT Tomato";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ011576; CAA09697.1; .  
KW RNA-directed RNA polymerase.  
SQ SEQUENCE 1116 AA; 127553 MW; 8D9B89DA826D31DB CRC64;  
Query Match 3.8%; Score 42; DB 10; Length 1116;  
Best Local Similarity 100.0%; Pred. No. 2.1e-36;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 979 EYDNKLGNLMDYGIKTEAEILSGGIMKASKTFDRRKDAEAI 1020  
DB 981 EYDNKLGNLMDYGIKTEAEILSGGIMKASKTFDRRKDAEAI 1022  
RESULT 3  
Q9ZRX3  
ID Q9ZRX3 PRELIMINARY; PRT: 775 AA.  
AC Q9ZRX3;  
DT 01-MAY-1999 (TREMELrel. 10, Created)  
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)  
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)  
DE RNA-directed RNA polymerase (fragment).  
GN RDRP.  
OS Petunia hybrida (Petunia).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Petunia.  
OX NCBI\_TaxID=4102;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LEAF;  
RA Schiebel W., Pellissier T., Riedel L., Thalmeir S., Schiebel R.,  
RA Kempe D., Lottspeich F., Saenger H.L., Wassenegeger M.;  
RT "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from  
RT Tomato";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ011979; CAA09896.1; .  
KW RNA-directed RNA polymerase.  
FT NON\_TER 1  
FT NON\_TER 775  
SQ SEQUENCE 775 AA; 88778 MW; DE43FE50C6A45B93 CRC64;  
Query Match 3.4%; Score 38; DB 10; Length 775;  
Best Local Similarity 100.0%; Pred. No. 3.9e-32;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 775 ALHMYDVCVFPQKGRPHNEGSGDLDGDIYFVCWD 812

Db 468 ALHMVDCVFPQKRPHPNECSGSLDGDIFYVCWD 505  
|||||

## RESULT 4

Q92RX5 PRELIMINARY; PRT; 513 AA.

AC Q92RX5; 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE RNA-directed RNA polymerase (Fragment).

GN RDRP.

OS Arabidopsis thaliana (Mouse-ear cross).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV, COLUMBIA; TISSUE=LEAF;

RA Schiebel W., Pelissier T., Riedel L., Thalmair S., Schiebel R.,

RA Kempe D., Lottspeich F., Saenger H.L., Wassenecker M.;

RT "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from

RT Tomato.";

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ011977; CAA09894.1;

KW RNA-directed RNA polymerase.

FT NON\_TER 1

FT NON\_TER 513

SQ SEQUENCE 513 AA; 57802 MW; 924FDA2B70EAD0E CRC64;

## Query Match

Best Local Similarity 2.1%; Score 23; DB 10; Length 513;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 791 RHPNECSGSLDGDIFYVCWDQ 813

|||||

Db 309 RHPNECSGSLDGDIFYVCWDQ 331

## RESULT 5

Q91QV2 PRELIMINARY; PRT; 1107 AA.

AC Q91QV2; 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE F10B6.19.

OS Arabidopsis thaliana (Mouse-ear cross).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Chao Q., Shinn P., Dunn P., Buehler E., Kahn S., Kim C., Walker M.,

RA Williams S., Altafi H., Araujo R., Conn L., Conway A.B., Gonzalez A.,

RA Hansen N.F., Huizar L., Kremenetskaia I., Lenz C., Li J., Liu S.,

RA Lutos S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V., Yu G.,

RA Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;

RT "Genomic sequence for Arabidopsis thaliana BAC F10B6 from chromosome

RT I.";

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Ecker J.R.;

RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Ecker J.R.;

RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

RN [4]

## RP SEQUENCE FROM N.A.

RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chlou J., Choi E.,  
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,  
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,  
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
RA Theologis A., Ecker J.;

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC006917; AAF79241.1;

SQ SEQUENCE 1107 AA; 126214 MW; F8FF97691FF682187 CRC64;

## Query Match

Best Local Similarity 2.1%; Score 23; DB 10; Length 1107;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 791 RHPNECSGSLDGDIFYVCWDQ 813

|||||

Db 787 RHPNECSGSLDGDIFYVCWDQ 809

## RESULT 6

Q92RX4 PRELIMINARY; PRT; 107 AA.

AC Q92RX4;

DT 01-MAY-1999 (TREMBlrel. 10, Created)

DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE RNA-directed RNA polymerase (Fragment).

GN RDRP.

OS Triticum aestivum (Wheat).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;

OC Triticeae; Triticum.

OX NCBI\_TaxID=4565;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=LEAF;

RA Schiebel W., Pelissier T., Riedel L., Thalmair S., Schiebel R.,

RA Kempe D., Lottspeich F., Saenger H.L., Wassenecker M.;

RT "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from

RT Tomato.";

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ011978; CAA09895.1;

KW RNA-directed RNA polymerase.

FT NON\_TER 1

FT NON\_TER 107

SQ SEQUENCE 107 AA; 11908 MW; D74821910CEA964E CRC64;

## Query Match

Best Local Similarity 1.6%; Score 18; DB 10; Length 107;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 792 PHPNECSGSLDGDIFYV 809

|||||

Db 1 PHPNECSGSLDGDIFYV 18

## RESULT 7

O82504 PRELIMINARY; PRT; 1133 AA.

AC O82504;

DT 01-NOV-1998 (TREMBlrel. 08, Created)

DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)

DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE F2P3.11 protein (Putative RNA-directed RNA polymerase).

GN F2P3.11 OR T2B4.110 OR AT4G11130.

OS Arabidopsis thaliana (Mouse-ear cross).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA WASHU;
RT "The A. thaliana Genome Sequencing Project.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Strong C., Graves T., Duckels G.;
RT "The sequence of A. thaliana F2P3.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Waterston R.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE FROM N.A.
RA Bevan M., Pohl T., Weizenegger T., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Schueller C.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
[6]
RN SEQUENCE FROM N.A.
RA Pohl T., Weizenegger T., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[7]
RN SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF080120; AAC35535.1; -
DR EMBL; AL049876; CAB43048.1; -
DR EMBL; AL161531; CAB81214.1; -
KW Nucleotidyltransferase; RNA-directed RNA polymerase; Transferase.
SQ SEQUENCE 1133 AA; 19323 MW; 32B72CE4E29B50B9 CRC64;

Query Match 1.3%; Score 14; DB 10; Length 1133;
Best Local Similarity 100.0%; Pred. No. 1.7e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 549 PSAFOIRYGGYKGV 562
Db 578 PSAFOIRYGGYKGV 591

RESULT 8
Q9SG02 PRELIMINARY; PRT; 1196 AA.
AC Q9SG02;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE Putative RNA-directed RNA polymerase.
GN T1G12.20 OR T9C5.95.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,
RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome 1 BAC T1G12 genomic sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RA Rieger M., Gabel C., Mueller-Auer S., Schaefer M., Zipp M.,

```

---

```

RA Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20306668; PubMed=10850495;
RA Mourrain P., Beclin C., Elmayan T., Feuerbach F., Godon C.,
RA Morel J.-B., Jovette D., Lacombe A.-M., Nikic S., Picault N.,
RA Remoue K., Sanial M., Vo T.-A., Vaucheret H.;
RT "Arabidopsis SGS2 and SGS3 genes are required for Posttranscriptional
RT Gene Silencing and Natural Virus Resistance.";
RL Cell 101:533-542(2000).
DR EMBL; AC012329; AAG52184.1; -
DR EMBL; AL132964; CAB71285.1; -
DR EMBL; AF239718; AAF73959.1; -
KW RNA-directed RNA polymerase.
SQ SEQUENCE 1196 AA; 136927 MW; 812DEB9EDC82C8F CRC64;

Query Match 1.1%; Score 12; DB 10; Length 1196;
Best Local Similarity 100.0%; Pred. No. 0.0028;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1047 KASAWYHVTYHP 1058
Db 1124 KASAWYHVTYHP 1135

RESULT 9
Q9LKP0 PRELIMINARY; PRT; 1196 AA.
AC Q9LKP0;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE RNA-dependent RNA polymerase.
GN SDE1
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20306669; PubMed=10850496;
RA Dalmay T., Hamilton A., Rudd S., Angell S., Baulcombe D.C.;
RT "An RNA-dependent RNA polymerase gene in Arabidopsis is required for
RT posttranscriptional gene silencing mediated by a transgene but not by
RT a virus.";
RL Cell 101:543-553(2000).
DR EMBL; AF268093; AAF74208.1; -
KW RNA-directed RNA polymerase.
SQ SEQUENCE 1196 AA; 136747 MW; C1B1961CDC11F708 CRC64;

Query Match 1.1%; Score 12; DB 10; Length 1196;
Best Local Similarity 100.0%; Pred. No. 0.0028;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1047 KASAWYHVTYHP 1058
Db 1124 KASAWYHVTYHP 1135

RESULT 10
Q95WU3 PRELIMINARY; PRT; 1377 AA.
AC Q95WU3;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE RNA-dependent RNA polymerase.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;

```

```

RN  [1]
RP  SEQUENCE FROM N.A.
RA  Lujan H.D.;
RT  "Identification of an RNA-dependent RNA-polymerase gene in giardia
RL  intestinalis.";
RL  Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AF293414; AAK97084.1; -.
DR  InterPro; IPR001865; RIBOSOMAL_S2.
DR  PROSITE; PS00962; RIBOSOMAL_S2_1; UNKNOWN.1.
SQ  SEQUENCE 1377 AA; 155267 MW; 7453D8646261A0B9 CRC64;

Query Match          0.9%; Score 10; DB 5; Length 1377;
Best Local Similarity 100.0%; Pred. No. 0.51;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  798 GSGSLDGDYI 807
    |||||
DB  680 GSGSLDGDYI 689

RESULT 11
ID  026600 PRELIMINARY; PRT; 583 AA.
AC  026600;
DT  01-JAN-1998 (TREMBLrel. 05, Created)
DT  01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT  01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE  Hypothetical protein MTH500.
GN  MTH500.
OS  Methanobacterium thermoautotrophicum.
OC  Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC  Methanobacteriaceae; Methanothermobacter.
OX  NCBI_TaxID=187420;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=98037514; PubMed=9371463;
RA  Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,
RA  Aldredge T., Bashiraden R., Blakely D., Cook R., Gilbert K.,
RA  Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA  Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA  Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Brabakhar S.,
RA  McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA  Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT  "Complete genome sequence of Methanobacterium thermoautotrophicum
RT  deltaH: functional analysis and comparative genomics.";
RL  J. Bacteriol. 179:7135-7155 (1997).
DR  EMBL; AE000833; AAB85006.1; -.
DR  InterPro; IPR001365; A/AMP_deaminase.
DR  PROSITE; PS00485; A_DEAMINASE; UNKNOWN.1.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 583 AA; 66088 MW; A267B146F0EA6B7A CRC64;

Query Match          0.8%; Score 9; DB 17; Length 583;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  345 SPSITLDDG 353
    |||||
DB  221 SPSITLDDG 229

RESULT 12
QYXYP1 PRELIMINARY; PRT; 1069 AA.
ID  Q9XYP1
AC  Q9XYP1;
DT  01-NOV-1999 (TREMBLrel. 12, Created)
DT  01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT  01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE  Dosa protein.
GN  DOSA.
OS  Dictyostellium discoideum (Slime mold); Dictyostellida; Dictyostelium.
OC  Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.

```

```

OX  NCBI_TaxID=44689;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Postlethwait P., Steck T.L.;
RT  "Dosa, A Gene Related To Hypo-Osmotic Stress Response In Dictyostellium
RL  discoideum.";
RL  Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AF117611; AAD29638.1; -.
DR  EMBL; AF117611; AAD29638.1; -.
SQ  SEQUENCE 1069 AA; 122832 MW; BE6815D3B69A1C1C CRC64;

Query Match          0.8%; Score 9; DB 5; Length 1069;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  797 CSGSLDGDG 805
    |||||
DB  274 CSGSLDGDG 282

RESULT 13
ID  09C162 PRELIMINARY; PRT; 1638 AA.
AC  09C162;
DT  01-JUN-2001 (TREMBLrel. 17, Created)
DT  01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT  01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE  Suppressor of ascus dominance.
GN  SAD-1.
OS  Neurospora crassa.
OC  Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC  Sordariales; Sordariaceae; Neurospora.
OX  NCBI_TaxID=5141;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Shiu P.K.T., Raju N.B., Metzberg R.L.;
RT  "Meiotic Silencing of Unpaired DNA.";
RL  Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AY029284; AAK31733.1; -.
DR  InterPro; IPR000504; RNA_rec_mot.
DR  Pfam; PF00076; rrm; 1.
SQ  SEQUENCE 1638 AA; 184233 MW; 7BC58EA66373D9BE CRC64;

Query Match          0.8%; Score 9; DB 3; Length 1638;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  485 AKYAARLQG 493
    |||||
DB  653 AKYAARLQG 661

RESULT 14
ID  039857 PRELIMINARY; PRT; 198 AA.
AC  039857;
DT  01-JAN-1998 (TREMBLrel. 05, Created)
DT  01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT  01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE  Coat protein.
OS  grapevine leafroll-associated virus 2.
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC  Closterovirus.
OX  NCBI_TaxID=64003;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Abou-Ghanem N.;
RT  "The nucleotide sequence of the 3' terminal region of grapevine
RL  leafroll associated closterovirus 2.";
RL  Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR  EMBL; Y14131; CAA74566.1; -.
DR  InterPro; IPR002679; Closter_coat.
DR  Pfam; PF01785; Closter_coat; 1.
SQ  SEQUENCE 198 AA; 21630 MW; 15D631E4088D3F01 CRC64;

```

Query Match 0.7%; Score 8; DB 12; Length 198;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 LLSAEVVK 20  
| | | | | | | |  
Db 27 LLSAEVVK 34

RESULT 15  
Q12044 Q12044 PRELIMINARY; PRT; 265 AA.  
AC Q12044;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
DE Chromosome XV reading frame ORF YOR220W.  
GN YOR220W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Boyer J., Fairhead C., Gaillon L., Galisson F., Michaux G.,  
RA Thierry A., Dujon B.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA MIPS;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-FY1679;  
RA Galisson F., Dujon B.;  
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; 275128; CAA95438.1; -;  
DR EMBL; X92441; CAA63183.1; -;  
DR SGD; S0005746; YOR220W.  
SQ SEQUENCE 265 AA; 29255 MW; 4A184C5C1218A517 CRC64;

Query Match 0.7%; Score 8; DB 3; Length 265;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 PPKSPSIT 349  
| | | | | | | |  
Db 252 PPKSPSIT 259

RESULT 16  
Q92Y23 Q92Y23 PRELIMINARY; PRT; 286 AA.  
AC Q92Y23;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)  
DE Hypothetical protein RAL065.  
GN RAL065 OR SWA1945.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OG Plasmid pSymA (megaplasmid 1).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=21396509; PubMed=11481432;  
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,  
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Guzy J.,  
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn M.L.,  
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,

RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;  
RT "Nucleotide sequence and predicted functions of the entire  
RT Sinorhizobium meliloti pSymA megaplasmid.",  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).  
DR EMBL; AE007292; AAK65723.1; -;  
DR InterPro; IPR000620; DUF6.  
DR Pfam; PF00892; DUF6; 1.  
KW Plasmid; Hypothetical protein; Complete proteome.  
SQ SEQUENCE 286 AA; 30992 MW; 4D8CC747B026FDA4 CRC64;

Query Match 0.7%; Score 8; DB 16; Length 286;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 683 SKLLDLRT 690  
| | | | | | | |  
Db 251 SKLLDLRT 258

RESULT 17  
Q97ES6 Q97ES6 PRELIMINARY; PRT; 352 AA.  
AC Q97ES6;  
DT 01-OCT-2001 (TREMBLrel. 18, Created)  
DT 01-OCT-2001 (T-EMBLrel. 18, Last sequence update)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
DE Histidinol-phosphate aminotransferase.  
GN CAC3031.  
OS Clostridium acetobutylicum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
OC Clostridiales; Clostridiaceae; Clostridium.  
OX NCBI\_TaxID=1468;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
RX MEDLINE=21359325; PubMed=11466286;  
RA Neelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
RA Tatusov R.I., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
RA Bennett G.N., Koonin E.V., Smith D.R.;  
RT "Genome sequence and comparative analysis of the solvent-producing  
RT bacterium Clostridium acetobutylicum";  
RL J. Bacteriol. 183:4823-4838(2001).  
DR EMBL; AE007800; AAK80971.1; -;  
DR InterPro; IPR002106; AATRNA\_ligaseI.  
DR InterPro; IPR004839; Aminotransf1/2.  
DR InterPro; IPR001917; NHtransf\_2.  
DR Pfam; PF00155; aminotran\_1\_2; 1.  
DR TIGRfams; TIGR01141; hisC; 1.  
DR PROSITE; PS00599; AA\_TRANSFER\_CLASS\_2; UNKNOWN\_1.  
DR PROSITE; PS00339; AA\_TRNA\_LIGASE\_II\_2; UNKNOWN\_1.  
KW Transferase; Aminotransferase; Complete proteome.  
SQ SEQUENCE 352 AA; 40220 MW; 3375021EBEC6BC2E CRC64;

Query Match 0.7%; Score 8; DB 16; Length 352;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 388 IDNFLRVS 395  
| | | | | | | |  
Db 326 IDNFLRVS 333

RESULT 18  
Q95ZS5 Q95ZS5 PRELIMINARY; PRT; 444 AA.  
AC Q95ZS5;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
DE F56A8.3a protein.  
GN F56A8.3a.  
OS Caenorhabditis elegans.



OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OX Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MCMurray A.A.;  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
RT investigating biology.";  
RL Science 282:2012-2018(1998).  
DR EMBL; Z83230; CAC42304.1; -;  
DR InterPro; IPR001611; LRR.  
DR Pfam; PF00560; LRR; 3.  
SQ SEQUENCE 444 AA; 49409 MW; 72E90B3851DA45DB CRC64;  
  
Query Match 0.7%; Score 8; DB 5; Length 444;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 637 DSLKAE 644  
Db 156 DSLKAE 163  
  
RESULT 19  
Q9CU85 PRELIMINARY; PRT; 490 AA.  
AC Q9CU85;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DE 5430414H02Rik protein (Fragment).  
GN PEX1 OR 5430414H02Rik.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=HEAD;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,  
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
DR EMBL; AK017303; BAB30684.1; -;  
DR MGD; MGI:1918632; PEX1.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003959; AAA\_ATPase\_cent.  
DR InterPro; IPR003960; AAA\_sub.  
DR Pfam; PF00004; AAA; 1.  
DR SMART; SM00382; AAA; 1.  
DR PROSITE; PS00674; AAA; UNKNOWN\_1.

KW ATP-binding.  
FT NON\_TER 1  
SQ SEQUENCE 490 AA; 53417 MW; 01DDC00B35C5BDA6 CRC64;  
  
Query Match 0.7%; Score 8; DB 11; Length 490;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 68 RLYFGSSY 75  
Db 334 RLYFGSSY 341  
  
RESULT 20  
Q9SP43 PRELIMINARY; PRT; 515 AA.  
AC Q9SP43;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DE 01-MAR-2001 (TrEMBLrel. 20, Last annotation update)  
DE Glucose-1-phosphate adenylyltransferase (EC 2.7.7.27) (ADP-glucose  
DE synthase) (ADP-glucose pyrophosphorylase).  
GN AGPS.  
OS Citrus unshiu (Satsuma orange).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosid II; Sapindales; Rutaceae; Citrus.  
OX NCBI\_TaxID=55188;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21210979; PubMed=11311948;  
RA Kim I.-J., Noh S.-J., Lee B.-H., Jo J., Kim Y.-S., Chung W.-I.;  
RT "Molecular characterization of cDNA clones for ADP-glucose  
RT pyrophosphorylase from Citrus.";  
RL Biochim. Biophys. Acta 1518:324-328(2001).  
CC -1- CATALYTIC ACTIVITY: ATP + ALPHA-D-GLUCOSE 1-PHOSPHATE =  
CC DIPHOSPHATE + ADP-GLUCOSE.  
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL AND PLANTS GLUCOSE-1-  
CC PHOSPHATE ADENYLYLTRANSFERASE FAMILY.  
DR EMBL; AF184597; AAD56041.1; -;  
DR InterPro; IPR001825; NTP\_transferase.  
DR Pfam; PF00483; NTP\_transferase; 1.  
DR PROSITE; PS00809; ADP\_GLC\_PYROPHOSPH\_1; 1.  
DR PROSITE; PS00809; ADP\_GLC\_PYROPHOSPH\_2; 1.  
DR PROSITE; PS00810; ADP\_GLC\_PYROPHOSPH\_3; 1.  
KW Glycogen biosynthesis; Nucleotidyltransferase; Transferase.  
SQ SEQUENCE 515 AA; 56814 MW; 925CED3EC3F44258 CRC64;  
  
Query Match 0.7%; Score 8; DB 10; Length 515;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 445 LAFSSSQL 452  
Db 34 LAFSSSQL 41  
  
RESULT 21  
Q9DG87 PRELIMINARY; PRT; 593 AA.  
AC Q9DG87;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Glutamic acid decarboxylase isoform 65 (EC 4.1.1.15).  
GN GAD65.  
OS Carassius auratus (Goldfish).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Carassius.  
OX NCBI\_TaxID=7957;  
RN [1]  
RP SEQUENCE FROM N.A.

RC TISSUE-BRAIN, AND PITUITARY;  
RX MEDLINE=21103894; PubMed=11181814;  
RA Bosma P.T., Blazquez M., Fraser E.J., Schulz R.W., Docherty K.,  
RT Trudeau V.L.;  
RT "Sex steroid regulation of glutamate decarboxylase mRNA expression in  
RT goldfish brain is sexually dimorphic.";  
RL J. Neurochem. 76:945-956(2001)  
CC -|- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
CC -|- SIMILARITY: BELONGS TO GROUP II DECARBOXYLASES (DDC, GAD, HDC AND  
CC TYRDC).  
DR EMBL; AF045594; AAG33931.1; -.  
DR InterPro; IPR002129; Pyridoxal\_dec.  
DR Pfam; PF00282; Pyridoxal\_dec. 1.  
DR PROSITE; PS00392; DDC\_GAD\_HDC\_YDC; 1.  
KW Decarboxylase; Lyase; Pyridoxal phosphate.  
SQ SEQUENCE 583 AA; 65375 MW; 8C43E1A264D95001 CRC64;  
  
Query Match 0.7%; Score 8; DB 13; Length 583;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 404 LYSTDLLP 411  
DB 90 LYSTDLLP 97  
|||||  
  
RESULT 22  
Q9IAT4 PRELIMINARY; PRT; 583 AA.  
ID Q9IAT4  
AC Q9IAT4;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Glutamic acid decarboxylase isoform 65 (EC 4.1.1.15).  
GN GAD65.  
OS Carassius auratus (Goldfish).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Carassius.  
OX NCBI\_TaxID=7957;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RA Bosma P.T., Blazquez M., Docherty K., Trudeau V.L.;  
RT "Cloning and expression of multiple forms of glutamate decarboxylase  
RT from goldfish.";  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
CC -|- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
CC -|- SIMILARITY: BELONGS TO GROUP II DECARBOXYLASES (DDC, GAD, HDC AND  
CC TYRDC).  
DR EMBL; AF149832; AAF73186.1; -.  
DR InterPro; IPR002129; Pyridoxal\_dec.  
DR Pfam; PF00282; Pyridoxal\_dec. 1.  
DR PROSITE; PS00392; DDC\_GAD\_HDC\_YDC; 1.  
KW Decarboxylase; Lyase; Pyridoxal phosphate.  
SQ SEQUENCE 583 AA; 65361 MW; 9C4326BFD2023DB6 CRC64;  
  
Query Match 0.7%; Score 8; DB 13; Length 583;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 404 LYSTDLLP 411  
DB 90 LYSTDLLP 97  
|||||  
  
RESULT 23  
Q9Y232 PRELIMINARY; PRT; 598 AA.  
ID Q9Y232  
AC Q9Y232;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE Testis-specific CHROMODOMAIN Y-like protein.  
GN CDYL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-TESTIS;  
RX MEDLINE=99206615; PubMed=10192397;  
RA Lahn B.T., Page D.C.;  
RT "Retroposition of autosomal mRNA yielded testis-specific gene family  
RT on human Y chromosome.";  
RL Nat. Genet. 21:429-433(1999).  
CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -|- SIMILARITY: CONTAINS 1 "CHROMO" DOMAIN.  
DR EMBL; AF081259; AAD22735.1; -.  
DR EMBL; AF081258; AAD22734.1; -.  
DR HSSP; P23197; IAP0.  
DR InterPro; IPR000953; Chromo.  
DR Pfam; PF00385; Chromo; 1.  
DR Pfam; PF00378; ECH; 1.  
DR PRINTS; PR00504; CHROMODOMAIN.  
DR SMART; SM00298; CHROMO; 1.  
DR PROSITE; PS00598; CHROMO\_1; 1.  
DR PROSITE; PS00013; CHROMO\_2; 1.  
KW Nuclear protein.  
SQ SEQUENCE 598 AA; 66419 MW; 503D20E04BDB9C90 CRC64;  
  
Query Match 0.7%; Score 8; DB 4; Length 598;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 739 SRSTNSNF 746  
DB 138 SRSTNSNF 145  
|||||  
  
RESULT 24  
Q96YH3 PRELIMINARY; PRT; 635 AA.  
ID Q96YH3  
AC Q96YH3;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Putative glutamate synthase large subunit.  
GN ST2198.  
OS Sulfolobus tokodaii.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.  
OX NCBI\_TaxID=111955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JCM 10545 / 7;  
RX PubMed=11572479;  
RA Karabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
RA Sakine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,  
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
RA Yoshizawa T., Tanaka T., Kudoh T., Yamazaki J., Kushida N., Oguchi A.,  
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
RA Oshima T., Kikuchi H.;  
RT "Complete genome sequence of an aerobic thermoacidophilic  
RT Crenarchaeon, Sulfolobus tokodaii strain 7.";  
RL DNA Res. 8:123-140(2001).  
DR EMBL; AP000989; BAB67304.1; -.  
DR InterPro; IPR003009; FMN\_enzyme.  
DR InterPro; IPR002932; Glu\_synthase.  
DR Pfam; PF01645; Glu\_synthase; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 635 AA; 68139 MW; 8D6BEE204F66812F CRC64;  
  
Query Match 0.7%; Score 8; DB 17; Length 635;

```
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 498 SRETLSVL 505
DB 319 SRETLSVL 326
|||||||

RESULT 25
O65482 PRELIMINARY; PRT; 830 AA.
ID O65482;
AC O65482;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Serine/threonine kinase - like protein (Serine/threonine kinase-like
protein).
DE
GN F21P8.200 OR F16G20.10 OR A74G23310.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevar M., Massenet O., Clabault G., Quigley F., Mache R., Bancroft I.,
RA Mewes H.W., Mayer K., Schueller C.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 618-830 FROM N.A.
RA Watson M.D., Gibbons T., Bartley J., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Schueller C., Bevan M.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA Massenet O., Clabault G., Quigley F., Mache R., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE OF 617-830 FROM N.A.
RA Watson M.D., Gibbons T., Bartley J., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AL022347; CAA18478.1; -
DR EMBL; AL031326; CAA20452.1; -
DR EMBL; AL161559; CAB79286.1; -
DR HSP; P12931; 1FMK
DR InterPro; IPR002902; DUF26.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001759; Pentaxin.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR004040; STY_pkinase.
DR Pfam; PF01657; DUF26; 3.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00221; STYK; 1.
DR PROSITE; PS00289; PENTAXIN; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 830 AA; 93315 MW; C1785EEC50E482EF CRC64;

Query Match 0.7%; Score 8; DB 10; Length 830;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 606 LITLSTL 613
```

```
Db 162 LITLSTL 169
|||||||

RESULT 26
O96064 PRELIMINARY; PRT; 864 AA.
ID O96064;
AC O96064;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Paramyosin.
OS Mytilus galloprovincialis (Mediterranean mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=29158;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ANTERIOR BYSSUS RETRACTOR MUSCLE;
RX MEDLINE=200730661; PubMed=10607366;
RA Watabe S., Iwasaki K., Funabara D., Hirayama Y., Nakaya M.,
RA Kikuchi K.;
RT "Complete amino acid sequence of Mytilus anterior byssus retractor
paramyosin and its putative phosphorylation site.";
RL J. Exp. Zool. 286:24-35(2000).
DR EMBL; AB016070; BAA36517.1; -
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR004827; TE_bzip.
DR Pfam; PF01576; Myosin_tail; 1.
DR SMART; SM00338; BRLZ; 1.
SQ SEQUENCE 864 AA; 99572 MW; EF7C8ED73A5742B0 CRC64;

Query Match 0.7%; Score 8; DB 5; Length 864;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1101 SSLRAQLS 1108
|||||||
DB 280 SSLRAQLS 287

RESULT 27
O53347 PRELIMINARY; PRT; 1101 AA.
ID O53347;
AC O53347;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Putative ATP-dependent DNA helicase (Helicase, Uvid/Rep family).
GN RV3201C OR MTV014.45C OR MT3295.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
```

RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL021646; CAA16666.1; -;  
 DR EMBL; AE007142; AAK47638.1; -;  
 DR HSSP; P56255; 1PJR.  
 DR TIGR; W3235; -;  
 DR Tuberculist; rv3201c; -;  
 DR InterPro; IPR000212; UvrD-helicase.  
 DR Pfam; PF00580; UvrD-helicase; 1.  
 KW Helicase; Complete proteome.  
 SQ SEQUENCE 1101 AA; 116687 MW; 269C6AARD3657412B CRC64;

Query Match 0.7%; Score 8; DB 16; Length 1101;  
 Best Local Similarity 100.0%; Pred. No. 67;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1021 SVAVRALR 1028

Db 374 SVAVRALR 381

RESULT 28

Q96S73 PRELIMINARY; PRT; 1226 AA.

ID Q96S73  
 AC Q96S73;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Pexlp-634del1690.  
 GN PEX1-634DEL690.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tamura S., Matsumoto N., Imamura A., Shimozawa N., Suzuki Y.,  
 RA Kondo N., Fujiki Y.;  
 RT "Phenotype-Genotype Relationships in Peroxisome Biogenesis Disorders  
 RT of PEX1-Defective Complementation Group 1 Are Defined by Pexlp-Pex6p  
 RT Interaction.";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB052090; BAB59061.1; -;  
 DR InterPro; IPR003959; AAA\_ATPase\_cent.  
 DR InterPro; IPR003960; AAA\_sub.  
 DR Pfam; PF00004; AAA; 2.  
 DR PROSITE; PS00674; AAA; UNKNOWN\_1.  
 SQ SEQUENCE 1226 AA; 136584 MW; 28125C8B785DA8E2 CRC64;

Query Match 0.7%; Score 8; DB 4; Length 1226;  
 Best Local Similarity 100.0%; Pred. No. 74;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 RLYFGSSY 75

Db 1070 RLYFGSSY 1077

RESULT 29

Q96S72 PRELIMINARY; PRT; 1283 AA.

ID Q96S72  
 AC Q96S72;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Pexlp1L664P.  
 GN PEX1L664P.  
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tamura S., Matsumoto N., Imamura A., Shimozawa N., Suzuki Y.,  
 RA Kondo N., Fujiki Y.;  
 RT "Phenotype-Genotype Relationships in Peroxisome Biogenesis Disorders  
 RT of PEX1-Defective Complementation Group 1 Are Defined by Pexlp-Pex6p  
 RT Interaction.";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB052091; BAB59062.1; -;  
 DR InterPro; IPR003959; AAA\_ATPase\_cent.  
 DR InterPro; IPR003960; AAA\_sub.  
 DR Pfam; PF00004; AAA; 2.  
 DR PROSITE; PS00674; AAA; UNKNOWN\_1.  
 SQ SEQUENCE 1283 AA; 142850 MW; C33AD8A68893D77A CRC64;

Query Match 0.7%; Score 8; DB 4; Length 1283;  
 Best Local Similarity 100.0%; Pred. No. 77;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 RLYFGSSY 75

Db 1127 RLYFGSSY 1134

RESULT 30

Q96S71 PRELIMINARY; PRT; 1283 AA.

ID Q96S71  
 AC Q96S71;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE PexlpG843D.  
 GN PEX1G843D.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tamura S., Matsumoto N., Imamura A., Shimozawa N., Suzuki Y.,  
 RA Kondo N., Fujiki Y.;  
 RT "Phenotype-Genotype Relationships in Peroxisome Biogenesis Disorders  
 RT of PEX1-Defective Complementation Group 1 Are Defined by Pexlp-Pex6p  
 RT Interaction.";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB052092; BAB59063.1; -;  
 DR InterPro; IPR003959; AAA\_ATPase\_cent.  
 DR InterPro; IPR003960; AAA\_sub.  
 DR Pfam; PF00004; AAA; 2.  
 DR PROSITE; PS00674; AAA; UNKNOWN\_1.  
 SQ SEQUENCE 1283 AA; 142924 MW; 9CF230C26F02AFD9 CRC64;

Query Match 0.7%; Score 8; DB 4; Length 1283;  
 Best Local Similarity 100.0%; Pred. No. 77;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 RLYFGSSY 75

Db 1127 RLYFGSSY 1134

RESULT 31

Q96S21 PRELIMINARY; PRT; 1344 AA.

ID Q96S21  
 AC Q96S21;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Chloroplast-resident DNA methyltransferase.  
 GN CRMET1.

```

OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CC683;
RX MEDLINE=21980658; PubMed=11983892;
RA Nishiyama R., Itoh M., Yamaguchi Y., Koizumi N., Sano H.;
RT "A chloroplast-resident DNA methyltransferase is responsible for
RT hypermethylation of chloroplast genes in Chlamydomonas maternal
RT gametes.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:5925-5930(2002).
KW EMBL; AB073989; BAB91073.1; -.
KW Transferase; Methyltransferase.
SQ SEQUENCE 1344 AA; 142361 MW; 96D3876E7EEB477F CRC64;

Query Match 0.7%; Score 8; DB 10; Length 1344;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1100 LSSLRAQL 1107
DB 150 LSSLRAQL 157
|||||

RESULT 32
Q9XW49 PRELIMINARY; PRT; 1564 AA.
AC Q9XW49;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Y53C10A.9 protein.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA White S.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AL033536; CAA22142.1; -.
DR InterPro; IPR003439; ABC.transportr.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00005; ABC.trans; 2.
DR ProDom; PD000006; ABC.transportr; 2.
DR PROSITE; PS00211; ABC.TRANSPORTER; UNKNOWN_2.
DR PROSITE; PS00636; DNAJ_1; UNKNOWN_1.
SQ SEQUENCE 1564 AA; 176701 MW; DC66A3E7ED8DC50F CRC64;

Query Match 0.7%; Score 8; DB 5; Length 1564;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 606 LITLSTL 613
DB 254 LITLSTL 261
|||||

RESULT 33
Q95ZG7 PRELIMINARY; PRT; 1883 AA.
AC Q95ZG7;
DT 01-DEC-2001 (TREMBlrel. 19, Created)

```

```

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative RNA dependent RNA polymerase (Fragment).
GN RPPA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Martens H., Novotny J., Oberstrass J., Steck T.L., Postlethwait P.,
RA Nellen W.;
RT "RNAi in Dictyostelium: developmental regulation and the role of RdRPs
RT and dsRNase.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ314909; CAC41974.1; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR001201; PAP_25A_core.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
KW ATP-binding; Helicase.
FT NON_TER 1883
SQ SEQUENCE 1883 AA; 217454 MW; 1B1D9B9738B15094 CRC64;

Query Match 0.7%; Score 8; DB 5; Length 1883;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 798 SGSDLDGD 805
DB 1353 SGSDLDGD 1360
|||||

RESULT 34
Q95ZG6 PRELIMINARY; PRT; 1964 AA.
AC Q95ZG6;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative RNA dependent RNA polymerase (Fragment).
GN RPPB.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Martens H., Novotny J., Oberstrass J., Steck T.L., Postlethwait P.,
RA Nellen W.;
RT "RNAi in Dictyostelium: developmental regulation and the role of RdRPs
RT and dsRNase.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ314910; CAC41975.1; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR001201; PAP_25A_core.
DR Pfam; PF00271; Helicase_C; 1.
KW ATP-binding; Helicase.
FT NON_TER 1
SQ SEQUENCE 1964 AA; 227413 MW; B83CA9098675E645 CRC64;

Query Match 0.7%; Score 8; DB 5; Length 1964;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 798 SGSDLDGD 805
DB 1177 SGSDLDGD 1184
|||||

```

```

RESULT 35
Q9ETK0
ID Q9ETK0 PRELIMINARY; PRT: 3229 AA.
AC Q9ETK0;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Putative methylase (or helicase).
GN ORF26.
OS Corynebacterium equi (Rhodococcus equi).
OG Plasmid PREAT701 (p33701), and Plasmid virulence.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=43767;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC33701; PLASMID-PREAT701 (P33701);
RA Takai S., Sekizaki T., Kakuda T., Nakamura M., Suzuki K., Ogino N.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC33701, AND 103; PLASMID-PREAT701 (P33701), AND VIRULENCE;
RX PubMed11083803;
RA Takai S., Hines S.A., Sekizaki T., Nicholson V.M., Alperin D.A.,
RA Osaki M., Takamatsu D., Nakamura M., Suzuki K., Ogino N., Kakuda T.,
RA Dan H., Prescott J.F.;
RT "DNA sequence and comparison of virulence plasmids from Rhodococcus
equi ATCC 33701 and 103.";
RL Infect. Immun. 68:6840-6847(2000).
DR EMBL: AP001204; BAB16635.1; -
DR EMBL: AF116907; AAG21729.1; -
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00271; helicase_C; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
KW ATP-binding; Helicase; Methyltransferase; Plasmid.
SQ SEQUENCE 3229 AA; 351003 MW; C1E03A6005424F1A CRC64;

Query Match 0.7%; Score 8; DB 2; Length 3229;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 952 VARRSYDA 959
Db 2704 VARRSYDA 2711
|||||||

RESULT 36
Q8Z6B7
ID Q8Z6B7 PRELIMINARY; PRT: 29 AA.
AC Q8Z6B7;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein STV1874a.
GN STV1874a.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677508;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagsels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,

Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CM18.";
RL Nature 413:848-852(2001).
DR EMBL: AL627271; CAD02107.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 29 AA; 3223 MW; 2A834B159BD06DB4 CRC64;

Query Match 0.6%; Score 7; DB 16; Length 29;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 990 YYGKTE 996
Db 8 YYGKTE 14
|||||||

RESULT 37
Q9GKJ9
ID Q9GKJ9 PRELIMINARY; PRT: 36 AA.
AC Q9GKJ9;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Prostatic acid phosphatase (Fragment).
GN ACPp.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Van Poucke M., Yerie M., Tuggle C., Chardon P., Van Zeveren A.,
RA Peelman L.J.;
RT "Integration of porcine chromosome 13 maps.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF222911; AAG41124.1; -
DR HSSP: P15309; 2HPA.
DR InterPro: IPR000560; HisAc.phosphatse.
DR Pfam: PF00328; acid_phosphat; 1.
FT NON_TER 1
FT NON_TER 36
SQ SEQUENCE 36 AA; 3943 MW; 50C57B8C770E1A10C CRC64;

Query Match 0.6%; Score 7; DB 6; Length 36;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 PPEGISI 268
Db 12 PPEGISI 18
|||||||

RESULT 38
Q9HB48
ID Q9HB48 PRELIMINARY; PRT: 43 AA.
AC Q9HB48;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE Hypothetical 5.4 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Qihan L., Chenhong D., Hongling Z.;
RT "Putative human lung cell protein.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF281279; AAG13933.1; -
KW Hypothetical protein.

```

```
SQ SEQUENCE 43 AA; 5388 MW; 0FB612B8E31EBA71 CRC64;
Query Match 0.68; Score 7; DB 4; Length 43;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 LKKEFFF 133
|||||
Db 21 LKKEFFF 27

RESULT 39
ID Q8YNP2 PRELIMINARY; PRT; 66 AA.
AC Q8YNP2;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein Asr4522.
GN ASR4522.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003596; BAB76221.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 66 AA; 7573 MW; AB4150A4459CAA9 CRC64;

Query Match 0.68; Score 7; DB 16; Length 66;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 260 VQPPEGI 266
|||||
Db 55 VQPPEGI 61

RESULT 40
ID Q9GUX8 PRELIMINARY; PRT; 70 AA.
AC Q9GUX8;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE Foot protein-3A.
GN MGFP-3A.
OS Mytilus galloprovincialis (Mediterranean mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=29158;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FOOT;
RA Inoue K.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=FOOT;
RX MEDLINE=96305382; PubMed=8706704;
RA Inoue K., Takeuchi Y., Miki D., Odo S., Harayama S., Waite H.;
RT "Cloning, sequencing and sites of expression of genes for the
RT hydroxyarginine-containing adhesive-plaque protein of the mussel
RT Mytilus galloprovincialis.";
```

```
RL Eur. J. Biochem. 239:172-176(1996).
DR EMBL; AB049579; BAB16314.1;
SQ SEQUENCE 70 AA; 8003 MW; A902B3E300F5B9EA CRC64;

Query Match 0.68; Score 7; DB 5; Length 70;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 555 RYGGYKG 561
|||||
Db 49 RYGGYKG 55

RESULT 41
ID Q9NAV2 PRELIMINARY; PRT; 72 AA.
AC Q9NAV2;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE Foot protein-3 precursor variant1.
GN MEFP3.
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=6550;
RN [1]
RP SEQUENCE FROM N.A.
RX Warner S.C., Waite J.H.;
RT "Expression of multiple forms of an adhesive plaque protein in an
RT individual mussel, Mytilus edulis.";
RL Mar. Biol. 134:729-734(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX Warner S.C., Waite J.H.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF286136; AAF89278.1;
SQ SEQUENCE 72 AA; 8295 MW; 7D42746A910D99 CRC64;

Query Match 0.68; Score 7; DB 5; Length 72;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 555 RYGGYKG 561
|||||
Db 52 RYGGYKG 58

RESULT 42
ID Q9NAU5 PRELIMINARY; PRT; 72 AA.
AC Q9NAU5;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE Foot protein-3 precursor variant10.
GN MEFP3.
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=6550;
RN [1]
RP SEQUENCE FROM N.A.
RX Warner S.C., Waite J.H.;
RT "Expression of multiple forms of an adhesive plaque protein in an
RT individual mussel, Mytilus edulis.";
RL Mar. Biol. 134:729-734(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX Warner S.C., Waite J.H.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF286147; AAF89289.1;
SQ SEQUENCE 72 AA; 8214 MW; 7600241D884D6BF4 CRC64;
```

Query Match 0.6%; Score 7; DB 5; Length 72;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 555 RYGGYKG 561  
 |||||  
 DB 52 RYGGYKG 58

RESULT 43  
 Q9NAU4 PRELIMINARY; PRT; 72 AA.  
 ID Q9NAU4  
 AC Q9NAU4  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE Foot protein-3 precursor variant11.  
 GN MEFP3.  
 OS Mytilus edulis (Blue mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;  
 OC Mytiloidea; Mytilidae; Mytilus.  
 OX NCBI\_TaxID=6550;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Warner S.C., Waite J.H.;  
 RT "Expression of multiple forms of an adhesive plaque protein in an  
 individual mussel, Mytilus edulis.";  
 RL Mar. Biol. 134:729-734(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Warner S.C., Waite J.H.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF286148; AAF89290.1;  
 SQ SEQUENCE 72 AA; 8249 MW; 3198746A89796D81 CRC64;

Query Match 0.6%; Score 7; DB 5; Length 72;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 555 RYGGYKG 561  
 |||||  
 DB 52 RYGGYKG 58

RESULT 44  
 Q9NAU3 PRELIMINARY; PRT; 72 AA.  
 ID Q9NAU3  
 AC Q9NAU3  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE Foot protein-3 precursor variant12.  
 GN MEFP3.  
 OS Mytilus edulis (Blue mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;  
 OC Mytiloidea; Mytilidae; Mytilus.  
 OX NCBI\_TaxID=6550;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Warner S.C., Waite J.H.;  
 RT "Expression of multiple forms of an adhesive plaque protein in an  
 individual mussel, Mytilus edulis.";  
 RL Mar. Biol. 134:729-734(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Warner S.C., Waite J.H.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF286149; AAF89291.1;  
 SQ SEQUENCE 72 AA; 8298 MW; 7602746A8D6D6BF4 CRC64;

Query Match 0.6%; Score 7; DB 5; Length 72;  
 Best Local Similarity 100.0%; Pred. No. 68;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 555 RYGGYKG 561  
 |||||  
 DB 52 RYGGYKG 58

RESULT 45  
 Q9NAU2 PRELIMINARY; PRT; 72 AA.  
 ID Q9NAU2  
 AC Q9NAU2  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE Foot protein-3 precursor variant13.  
 GN MEFP3.  
 OS Mytilus edulis (Blue mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;  
 OC Mytiloidea; Mytilidae; Mytilus.  
 OX NCBI\_TaxID=6550;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Warner S.C., Waite J.H.;  
 RT "Expression of multiple forms of an adhesive plaque protein in an  
 individual mussel, Mytilus edulis.";  
 RL Mar. Biol. 134:729-734(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Warner S.C., Waite J.H.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF286150; AAF89292.1;  
 SQ SEQUENCE 72 AA; 8261 MW; B2B1746A98989D90 CRC64;

Query Match 0.6%; Score 7; DB 5; Length 72;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 555 RYGGYKG 561  
 |||||  
 DB 52 RYGGYKG 58

RESULT 46  
 Q9NAU7 PRELIMINARY; PRT; 72 AA.  
 ID Q9NAU7  
 AC Q9NAU7  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE Foot protein-3 precursor variant19.  
 GN MEFP3.  
 OS Mytilus edulis (Blue mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;  
 OC Mytiloidea; Mytilidae; Mytilus.  
 OX NCBI\_TaxID=6550;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Warner S.C., Waite J.H.;  
 RT "Expression of multiple forms of an adhesive plaque protein in an  
 individual mussel, Mytilus edulis.";  
 RL Mar. Biol. 134:729-734(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Warner S.C., Waite J.H.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF286156; AAF89298.1;  
 SQ SEQUENCE 72 AA; 8327 MW; 6CB3DF69ED6D6BF4 CRC64;

Query Match 0.6%; Score 7; DB 5; Length 72;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 555 RYGGYKG 561



Db 52 RYGGK 58  
|||||||

## RESULT 47

Q9TTV1 PRELIMINARY; PRT; 78 AA.  
AC Q9TTV1  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE Breast basic conserved protein 1 (Fragment).  
OS Ovis aries (Sheep).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Caprinae; Ovis.  
OX NCBI\_TaxID=9940;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PULMONARY VEIN;  
RA Sander F.C., Raj J.U.;  
RT "Breast Basic Conserved Gene (BBCL) In Fetal Pulmonary Vascular Development";  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF192977; AAF03752.1;  
DR InterPro; IPR001380; Ribosomal\_L13E.  
DR Pfam; PF01294; Ribosomal\_L13e; 1.  
DR ProDom; PD004443; Ribosomal\_L13E; 1.  
FT NON\_TER 1  
FT NON\_TER 78  
SQ SEQUENCE 78 AA; 8924 MW; 54EBA54962C35E2D CRC64;

Query Match 0.6%; Score 7; DB 6; Length 78;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 VDPRRN 304  
|||||||

Db 67 VDPRRN 73

## RESULT 48

Q8ZLJ9 PRELIMINARY; PRT; 79 AA.  
AC Q8ZLJ9  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE putative membrane protein.  
GN SY4565.  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M., Baker S., Basham D., Brooks K., Chillingworth T., Connerton P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi Cn18";  
RL Nature 413:848-852(2001).  
DR EMBL; AJ627282; CAB09341.1;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 79 AA; 8818 MW; 747F99DE87ECB71C CRC64;

Query Match 0.6%; Score 7; DB 16; Length 79;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 456 SVMFAS 462  
|||||||

Db 73 SVMFAS 79

## RESULT 49

Q9NT43 PRELIMINARY; PRT; 89 AA.  
AC Q9NT43  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical 10.1 kDa protein (Fragment).  
GN DKFZP434P2119.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TESTIS;  
RA Otterwälder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL137543; CAB70801.1;  
DR InterPro; IPR004934; Tropomodulin.  
DR Pfam; PF03250; Tropomodulin; 1.  
KW Hypothetical protein.  
FT NON\_TER 1  
FT NON\_TER 89  
SQ SEQUENCE 89 AA; 10055 MW; 568C62F60D61D39A CRC64;

Query Match 0.6%; Score 7; DB 4; Length 89;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 652 ENTNILK 658  
|||||||

Db 48 ENTNILK 54

## RESULT 50

Q98I11 PRELIMINARY; PRT; 104 AA.  
AC Q98I11  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Hypothetical protein m112618.  
GN M112618.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S., Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T., Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti";  
RL DNA Res. 7:331-338(2000).  
DR EMBL; AP003000; BAB49705.1;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 104 AA; 10787 MW; 1743BC06301BAACB CRC64;

Query Match 0.6%; Score 7; DB 16; Length 104;

Best Local Similarity 100.0%; Pred. No. 95;  
Matches 7; Conservative 0; Mismatches 0;

0; Indels 0; Gaps 0;

Oy 347 SITLDDG 353

|||||

Db 60 SITLDDG 66

Search completed: November 6, 2002, 03:37:49  
Job time : 163 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:39:18 ; Search time 45 Seconds  
(without alignments)  
3298.693 Million cell updates/sec

Title: US-09-782-874-2

Perfect score: 5864

Sequence: 1 MGKTIQVGFPPYLLSAEVK.....RPVLNLSRLRAQLSHRLVVK 1114

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5864	100.0	1114	22	AAU10006
2	5864	100.0	1114	22	AAU10006
3	3643	62.1	1125	21	AAU28533
4	3140.5	53.6	1108	21	AAU28530
5	1621	27.6	1196	22	AAU67235
6	1591	27.1	1164	21	AAU28531
7	1430.5	24.4	1136	21	AAU28532
8	1182	20.2	218	22	AAU00898
9	1013.5	17.3	417	21	AAU28529
10	697.5	11.9	230	21	AAU28534

11	297	5.1	1402	21	AAU13956
12	205	3.5	60	21	AAU28537
13	197	3.4	90	21	AAU28536
14	167	2.8	111	21	AAU28535
15	146	2.5	2060	23	AAU20967
16	140	2.4	2029	23	ABU68683
17	129.5	2.2	849	23	ABU91193
18	128	2.2	157	21	AAU28528
19	126.5	2.2	2548	21	AAU53193
20	123	2.1	1498	22	ABU05469
21	120	2.0	996	22	ABU30156
22	117.5	2.0	908	22	ABU12986
23	117.5	2.0	1345	22	ABU13923
24	117	2.0	935	22	ABU10172
25	116.5	2.0	544	22	AAU01219
26	116.5	2.0	546	20	AAU06850
27	116.5	2.0	751	23	ABU26865
28	115	2.0	766	21	AAU14283
29	115	2.0	774	20	AAU37454
30	115	2.0	872	23	AAU071619
31	114.5	2.0	544	22	AAU01222
32	114.5	2.0	546	20	AAU06853
33	114.5	2.0	776	23	ABU4921
34	114	1.9	957	18	AAU20841
35	113.5	1.9	831	19	AAU54372
36	113.5	1.9	3226	22	ABU28408
37	113	1.9	763	21	AAU14284
38	112.5	1.9	742	22	ABU22122
39	112.5	1.9	988	19	AAU56107
40	112.5	1.9	1010	22	ABU20134
41	112.5	1.9	1184	22	AAU33733
42	112.5	1.9	1188	22	AAU36532
43	112	1.9	1198	23	ABU55148
44	111.5	1.9	544	22	AAU01225
45	111.5	1.9	546	20	AAU06856

#### ALIGNMENTS

#### RESULT 1

AAU10006  
ID AAU10006 standard; Protein; 1114 AA.

XX AAU10006;

XX 08-MAY-2002 (first entry)

XX Tomato RNA-directed RNA polymerase (RdRP) protein.

XX Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;  
XX cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;  
XX transgenic plant; transgenic animal; cancer; viral infection;  
XX immunoprecipitation; immunolocalisation; EC 2.7.7.48; gene therapy.

XX Lycopersicon esculentum.

XX US2001023067-A1.

XX 20-SEP-2001.

XX 08-FEB-2001; 2001US-0782874.

XX 05-MAR-1997; 97US-0811583.

XX (WASS/) WASENEGGER M.

XX (RIED/) RIEDEL L.

XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;

XX WPI; 2001-595798/67.

XX N-PSDB; AAS17837.

XX



/label= P431  
/note= "C-protein-specific antibody for  
immunisation of rabbits"

US6218142-B1.

17-APR-2001.

05-MAR-1997; 97US-0811583.

05-MAR-1997; 97US-0811583.

(WASS/) WASSENEGGER M.

(RIED/) RIEDEL L.

Wassenegger M, Riedel L, Schiebel W, Sanger HL;

WPI; 2001-289830/30.

N-PSDB; AAD04370.

New nucleic acid molecules encoding polypeptides with RNA-directed RNA polymerase enzymatic activity, useful in modulating gene expression in plants, humans and animals, as well as in plant cell/tissue cultures or plant breeding -

Claim 1; Column 33-40; 31pp; English.

The present sequence is a tomato C-protein having RNA-directed RNA polymerase (RdRP) activity. This protein is capable of RNA-directed RNA synthesis, thus using RNA as a template for synthesising complementary RNA molecules. RdRP nucleic acid is useful for modulating gene expression in plants, humans and animals. This may lead to various physiological, developmental and/or morphological changes. Transgenic plants containing RdRP nucleic acid is especially useful in plant cell or tissue cultures and in plant breeding. RdRP is useful in gene therapy, particularly for treating a disease that is caused by the undesirable expression or overexpression of a gene.

Sequence 1114 AA;

Query Match 100.0%; Score 5864; DB 22; Length 1114;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKTIQVGFPPYLLSAEYVKSFLKTYGTVCALVEYKQSGSRAFAKYQFADNISADK 60  
DB 1 MGKTIQVGFPPYLLSAEYVKSFLKTYGTVCALVEYKQSGSRAFAKYQFADNISADK 60  
QY 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDQMDGITLNFQCGISDDKFAVLGSTEVS 120  
DB 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDQMDGITLNFQCGISDDKFAVLGSTEVS 120  
QY 121 IQFGIGLKKFFFLSSGSADYKQLQSYENIWQVVLHRYPGQNAQFLLIQLFGAPRIYKRL 180  
DB 121 IQFGIGLKKFFFLSSGSADYKQLQSYENIWQVVLHRYPGQNAQFLLIQLFGAPRIYKRL 180  
QY 181 ENSCYSPFKETPDQWVRTTDFPSPWIGLSSSLCLOFRGRVRLPNPEESFFHYAERENNI 240  
DB 181 ENSCYSPFKETPDQWVRTTDFPSPWIGLSSSLCLOFRGRVRLPNPEESFFHYAERENNI 240  
QY 241 TLQTGFTFFYSQKALPNVPQPEGISIPYKILFKISLVLQHGCIQGPALNVYFFRLVDP 300  
DB 241 TLQTGFTFFYSQKALPNVPQPEGISIPYKILFKISLVLQHGCIQGPALNVYFFRLVDP 300  
QY 301 RRRNVACIEHALEKLYYKECCYDPVWLTQYDGYLKGQPKPSITLDDGLVYVRRV 360  
DB 301 RRRNVACIEHALEKLYYKECCYDPVWLTQYDGYLKGQPKPSITLDDGLVYVRRV 360  
QY 361 LVTPCKYVFCGPEVNSRVNRYNSDIDNLFVSVFDEWEKLYSTDLIPKASTGSGYR 420  
DB 361 LVTPCKYVFCGPEVNSRVNRYNSDIDNLFVSVFDEWEKLYSTDLIPKASTGSGYR 420  
QY 421 TNIYERILSTRKGFVIGDKKFFLAFFSSQLRDNVSMVFASRPGLTANDIRAWMGDFSQ 480

DB 421 TNIYERILSTRKGFVIGDKKFFLAFFSSQLRDNVSMVFASRPGLTANDIRAWMGDFSQ 480  
QY 481 IKWAKYAAARLQSGFSSRETLSVLREHEIVPDVKVHGTSTVFSGIGKISGDEFAHRVA 540  
DB 481 IKWAKYAAARLQSGFSSRETLSVLREHEIVPDVKVHGTSTVFSGIGKISGDEFAHRVA 540  
QY 541 SKGGLQYTPSAFQIRYGGYGVVDPDSMKLSLRKSMKSYESDNIKLDVLGWSKYQPC 600  
DB 541 SKGGLQYTPSAFQIRYGGYGVVDPDSMKLSLRKSMKSYESDNIKLDVLGWSKYQPC 600  
QY 601 YLNRQLITLLSTLGVKDEVLEQKQKAEVQDLDAILHDSLKAQEAELMSPGENTNLIKAM 660  
DB 601 YLNRQLITLLSTLGVKDEVLEQKQKAEVQDLDAILHDSLKAQEAELMSPGENTNLIKAM 660  
QY 661 LNCGYKPDAPFLSMMLQTFRASKLLDLTRSRIFIPNGRTMMGCLDESRTLEYGVQVYQ 720  
DB 661 LNCGYKPDAPFLSMMLQTFRASKLLDLTRSRIFIPNGRTMMGCLDESRTLEYGVQVYQ 720  
QY 721 FTGAGHGFSDDLHPNNSRSTNSNFILKGNVVAKNPCLHFGDIRVLKAVNVRALHHMV 780  
DB 721 FTGAGHGFSDDLHPNNSRSTNSNFILKGNVVAKNPCLHFGDIRVLKAVNVRALHHMV 780  
QY 781 DCVVPQKGRPHNPNBCSGDLGDGIYFYCWQDMLPPRQVOPMEYPPAPSIQLDHDVTI 840  
DB 781 DCVVPQKGRPHNPNBCSGDLGDGIYFYCWQDMLPPRQVOPMEYPPAPSIQLDHDVTI 840  
QY 841 EEVEEYFTNYIWNDSIGIITANAHVVFADREPDMAMSDPCKLAELFSIAVDFPKTGVPAE 900  
DB 841 EEVEEYFTNYIWNDSIGIITANAHVVFADREPDMAMSDPCKLAELFSIAVDFPKTGVPAE 900  
QY 901 IPSQLRPKEYPPFMDKPKDTSYISERVIGLFRKVKDKAPQASSIATFTRDVARSYDAD 960  
DB 901 IPSQLRPKEYPPFMDKPKDTSYISERVIGLFRKVKDKAPQASSIATFTRDVARSYDAD 960  
QY 961 MEVDGFEDYIDEAFDYKTEYDNKLGMLNDYIGIKTEAEILSGGIMKASKTFDRRKDAEAI 1020  
DB 961 MEVDGFEDYIDEAFDYKTEYDNKLGMLNDYIGIKTEAEILSGGIMKASKTFDRRKDAEAI 1020  
QY 1021 SVAVRALRKEARAWFKRRNDIDMLPKASAWHYHVIYHPTIYWCYNOGLKRAHFISFPNCV 1080  
DB 1021 SVAVRALRKEARAWFKRRNDIDMLPKASAWHYHVIYHPTIYWCYNOGLKRAHFISFPNCV 1080  
QY 1081 YDQLIQIKKDKARNRPVNLSSLRALQSLHRLVLK 1114  
DB 1081 YDQLIQIKKDKARNRPVNLSSLRALQSLHRLVLK 1114  
RESULT 3  
AAB28533  
ID AAB28533 standard; Protein; 1125 AA.  
XX  
AC AAB28533;  
DT 07-FEB-2001 (first entry)  
XX  
DE Soybean RNA-directed RNA polymerase, SEQ ID NO: 12.  
XX  
KW Soybean; plant; RNA-directed RNA polymerase; gene mapping; gene marker;  
KW plant virus resistance; plant breeding.  
XX  
OS Glycine max.  
XX  
PN WO2000060097-A1.  
XX  
PD 12-OCT-2000.  
XX  
PF 06-APR-2000; 2000WO-US09105.  
XX  
PR 07-APR-1999; 99US-0128094.  
XX  
PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
PA (PION-) PIONEER HI-BRED INT INC.









```
Db 386 VITPTKAYCMPPEVRSNRVIRHYSEVSDRFLRVTFMDGQMNLNSVNFSAQIVKDL 445
QY 410 LPKASTGSGVTRNIYERILSTLRKGVIGDKKFEFLAFSSSOLRDNVSMFASRPGLTAN 469
Db 446 M---SNSPLHKTIVYKRVKTELTGEGHMGGRKYSLAFSSNQLKDRSAWFAEDRTTVE 502
QY 470 DIRAMGDFSOIKNAVYAALRGSGSSRRTLSVLRHEI-EVIPDVKHVGTSYVFDGI 528
Db 503 TIRKMGGRFTS-KNVAKHAARMGOCFSTYATVVLQPHVEVNECLDEVEHNG--YIFSDGI 559
QY 529 GKISGDFAHRVASKGLQYT--PSAFOIRYGGYKGVGV--DPDSSMKLSLRKSMKYE 583
Db 560 GKITCDLALVAQK--LQTDNPPSAQIRYAGFAGFVSVWEGKNDGIRLSRPSMKFE 617
QY 584 SDNIKLDVLGMSKYQPCYNLRQLITLLSTLGVKDEVLQKQKEAVDQDLAILHDSLKAQE 643
Db 618 SNHTLEVSVTKFQPGFLNRQIITLLSSLVNPDPAIFAQMQEAMLSNLLNITLSDVAED 677
QY 644 ALELMSPGENTNILKAMLCNGYKPDAPFSLSMQLTFRASKLLDLRPSRIFINGRTMM 703
Db 678 -IVTASCAEQGTAAALMSAGISPGTEPHLKAMLLAIRSSQLLLEKTRIFVPGKRWLM 736
QY 704 GCLDESRTLEXQVFQVGTGAGHGBFSDDLHPFN---NSRSTNSGNFLLKGNVYVAKNPCL 760
Db 737 GCLDELGILEQGCQFIR---ASSPSLNNCLVYKGSRLSAANTNAETIL-GTIVAKNPCL 792
QY 761 HPGDTRVLKAVNRALHMHVDCVFPQKGRPHNECSGSLDGDITFVPCWQDMIPP-- 818
Db 793 HPGDVRILEAVDVPRLHLVCLVFPKKGPRHANEASGSLDGLDLYFVTWDELIPGK 852
QY 819 ROVQPMVEYPPAPSIQDLHDVTIEEVEEYFTNVIVNDSLGIITANAHVVPADREPDMAKSDP 878
Db 853 KSWNPMDSVPAPAKQLPRAVSOQHDIVGFLKKNVNEKLCPISNARHVHADSEYGADEK 912
QY 879 CKKLAEFLSIADVFPKTPGPAEIPSQLRPKEYPDMFKPKTSYISERVIGKLFKRYK-- 936
Db 913 CIQLAEALATAVDFFKTKIVSMPASTLRPKLYPDMFGKEDAIYSRSEKILGLYRSIQEA 972
QY 937 ---DKAPQASSTATFTRDVARSDADMEVDGFEDYIDEAFDYKTEYDNKLGNLMDYGI 993
Db 973 SDDLVPESTCTSN-----NLFPYDADMEVAGAADFLSSAMQCKCSYETQLNALLNQYV 1026
QY 994 KTEAEILSGIMKASKTFDRKRD--AEAISVAVRALRKEARAWFKR-----RNDIDD---- 1043
Db 1027 RTEAEIVTEHIWSLFPKYSRRQGDIKERLKNAYYALHKEERSIPESIVTDDTEISDDKS 1086
QY 1044 --MLPKASAWHYVTHPTYW-----GCYNOGLKRAHFISPPWCYVDQLQIK 1088
Db 1087 REYENKASAWYQVYHPE-VWQKSRMPFKSDCEDMPAR----LSFAWTAVEHLARIK 1138

RESULT 7
AAB28532
ID AAB28532 standard; Protein; 836 AA.
AC
AC AAB28532;
XT
DT 07-FEB-2001 (first entry)
DE
DE Rice RNA-directed RNA polymerase, SEQ ID NO: 10.
KW
KW Rice; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
KW plant virus resistance; plant breeding.
XX
XX
OS Oryza sativa.
XX
XX WO200060097-A1.
XX
XX 12-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US09105.
XX
```

```
PR 07-APR-1999; 99US-0128094.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
PA (PION-) PIONEER HI-BRED INT INC.
XX
XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;
XX
XX WI; 2000-679376/66.
DR N-PSDB; AAC63741.
XX
XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
XX for controlling gene expression and providing mechanisms to engineer
XX plant virus resistance
XX
XX Claim 10; Page 46-49; 62pp; English.
XX
XX The present sequence is part of a plant RNA-directed RNA polymerase
XX protein. Polynucleotides encoding RNA-directed RNA polymerases were
XX isolated from plant cDNA libraries. They are useful as probes for
XX genetically and physically mapping genes, and as markers for
XX traits linked to those genes. They are useful for controlling gene
XX expression and provide mechanisms to engineer plant virus resistance.
XX They are also useful for plant breeding to develop lines with desired
XX phenotypes.
XX
XX Sequence 836 AA;

Query Match 24.4%; Score 1430.5; DB 21; Length 836;
Best Local Similarity 40.9%; Pred. No. 7.2e-119;
Matches 336; Conservative 140; Mismatches 267; Indels 79; Gaps 22;

QY 343 PKSPSITLDGLVYVRVLTPTCKVYFCGPEVNVNRYSEDIDNLRVSVFVDEWE 402
Db 41 PKLCSKVGDDHAEVRSVITPTXAYCLPPEVRSNRYLRYHEVADRFLRVTFMDEGMQ 100
QY 403 KLXST-----DLLPKASTGSGVTRNIYERILSTLRKGVIGDKKFEFLAFSSSQ 451
Db 101 VLNNVLNSTAPIVKDLM---SNFQOKTIVYKVRMLLTGEGHMGGRKYSLAFSSNQ 157
QY 452 LRDSNVMSASRPLGTANDIRAMGDFSOIKNAVYAALRGSGSSRRTLSVLRHEIEV 511
Db 158 LRDSAMFFAEDRKTIVYARKWM-DGSQVR-MEEDARMGQCFSTYATVTRPDEVE 215
QY 512 IPDVKHVGTSYVSDGIGKISGDFAHRVASKCGLQYT---PSAFQIRYGGYKGVGV--D 566
Db 216 SFDDVWE-NEVIFS DGKITPDLAEVAER--LQTDNPPSAVQIRFAGFGVIAWQ 272
QY 567 PDSMKLSLRKSMKSYSDNIKLDVLGWSKYQPCYNLRQLITLLSTLGVKDEVLQKQKE 626
Db 273 HGDGTRFLRPSMRKFSNHLVGVVSWTKFQPGFLNRQIITLLSSLVNPDSTFWMQET 332
QY 627 AVDQDLAILHDSLKAQEALEIMSPGENTNILKAMLCNGYKPDAPFSLMMLQTRASKLL 686
Db 333 MSLNLLNLSDRDVAEVL-TTSCADDGNTAALMSAGFEPTPEPHLKAMLLAIRSAQLQ 391
QY 687 DLRTSRIFIPNGRTMMGCLDESRTLEYQGVFVQFTGAG-----HG-EFSDDLHPFN 738
Db 392 DLLEKARIFVPGKRWLMGCLDELGVLEQGCQFIRATVPVSLNSYFVKHGRFS----- 443
QY 739 SRSTNSNF-ILKGNVYVAKNPCLHPGDIRVLKAVNRALHMHVDCVFPQKGRPHNEC 797
Db 444 --STDKTEVILGTIVYAKNPCLHPGDIRVLEAVDVPRLHLVCLVFPQKGRPHANE 501
QY 798 SGSLDGDITFVPCWQDMIPP--RQVQPMYPPAPSTQLDHDVTIEVEYFTNYVND 855
Db 502 SGSLDGDITFVPCWQDMIPP--RQVQPMYPPAPSTQLDHDVTIEVEYFTNYVND 855
QY 856 LGIITANAHVVPADREPDMAKSDPCKKLAEFLSIADVFPKTPGPAEIPSQLRPKEYPDM 915
Db 562 LGRICNARHVHADSEYGADEKCIHLAEALTAATVDFPKTKLAIMPPLHKKVYDFPMG 621
QY 916 KPDKTSYISERVIGKLFKRYKDKAPQASSTATFTRDVARSS---YDADMEVDGFEDYIDE 972
XX
```

```
Db 622 KEDQSYKSEKILGRYSIQ-----EASNGDVVSQEVTPNDLPYDIDLEVPGASDFLAS 677
QY 973 AFYKTYDNKGLNDYIGIKTEAETISGIMKASKTFDRRKD--AEAISVAVRALRKE 1030
Db 678 AWCKCSYDAQLSALLSQYRTEAEVLVGHITFLVNSKKOGDIKDRLKTAYSALRKE 737
QY 1031 ARAWFK-----RRNDIDMLPKASAWHYHTYHTYGCYNQGL-----KRAH 1072
Db 738 FKSTFESIASDQCEIGDENKLLYEM--KASAWYQVTHPK-VWEKSRGILGPDGEEIPA 794
QY 1073 FIFPWCYVDLIQIK---KDKARNRPVLNLSRLAQLSHRL 1111
Db 795 SLSPWIPVDYLARIKLRCHGKVRVEGQKPVRLAAYISERI 836

RESULT 8
AAE00898
ID AAE00898 standard; Protein; 218 AA.
AC AAE00898;
XX
DT
DE 04-JUL-2001 (first entry)
XX
KW Tomato RNA-directed RNA polymerase 800 (RdRP800) protein.
XX
KW Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
KW transgenic plant; tissue culture; plant breeding; therapy; C-protein.
XX
OS Lycopersicon esculentum.
XX
PN US6218142-B1.
XX
PD 17-APR-2001.
XX
PF 05-MAR-1997; 97US-0811583.
XX
PR 05-MAR-1997; 97US-0811583.
PA (WASS/) WASENEGGER M.
PA (RIED/) RIEDEL L.
XX
PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX
XX WPI; 2001-289830/30.
XX
XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA
XX polymerase enzymatic activity, useful in modulating gene expression in
XX plants, humans and animals, as well as in plant cell/tissue cultures or
XX plant breeding
XX
XX Example 3; Column 39-42; 31pp; English.
XX
XX The present sequence is tomato RNA-directed RNA polymerase 800
XX (RdRP800) protein which corresponds to position 700-917 of C-protein.
XX C-protein is capable of RNA-directed RNA synthesis, thus using RNA as a
XX template for synthesizing complementary RNA molecules. RdRP nucleic acid
XX is useful for modulating gene expression in plants, humans and animals.
XX This may lead to various physiological, developmental and/or
XX morphological changes. Transgenic plants containing RdRP nucleic acid is
XX especially useful in plant cell or tissue cultures and in plant breeding.
XX RdRP is useful in gene therapy, particularly for treating a disease that
XX is caused by the undesirable expression or overexpression of a gene.
XX
XX Sequence 218 AA;
XX
XX Query Match 20.2%; Score 1182; DB 22; Length 218;
XX Best Local Similarity 100.0%; Pred. No. 1.6e-97;
XX Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 RTMGCIDSRLEYGVQVFOFTGAGHGEFSDDLHPFNNSRSTNSFILKGNVYVAKNPC 759
Db 1 RTMGCIDSRLEYGVQVFOFTGAGHGEFSDDLHPFNNSRSTNSFILKGNVYVAKNPC 60
```

```
QY 760 LHPGDIRVLKAVNRALHMHVDCVVPQKGRPHNECSGSLDGDIIYVCWQDMIPPR 819
Db 61 LHPGDIRVLKAVNRALHMHVDCVVPQKGRPHNECSGSLDGDIIYVCWQDMIPPR 120
QY 820 QVQPEYPPAPSIQLDHVDVTIEVEEYFTNYIVNDSLGIIANAHVVFADRPDMAMSDPC 879
Db 121 QVQPEYPPAPSIQLDHVDVTIEVEEYFTNYIVNDSLGIIANAHVVFADRPDMAMSDPC 180
QY 880 KKLAEFLSTAVDPFKTGVPAETPSQIRKPEYDFMDKP 917
Db 181 KKLAEFLSTAVDPFKTGVPAETPSQIRKPEYDFMDKP 218

RESULT 9
AAB28529
ID AAB28529 standard; Protein; 417 AA.
XX
AC AAB28529;
XX
DT 07-FEB-2001 (first entry)
XX
DE Maize partial RNA-directed RNA polymerase, SEQ ID NO: 4.
XX
KW Maize; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
KW plant virus resistance; plant breeding.
XX
OS Zea mays.
XX
PN WO200060097-A1.
XX
PD 12-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US09105.
XX
PR 07-APR-1999; 99US-0128094.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Odeil JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;
XX
XX WPI; 2000-679376/66.
XX N-PSDB; AAC63738.
XX
XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
XX for controlling gene expression and providing mechanisms to engineer
XX plant virus resistance
XX
XX Claim 10; Page 34-35; 62pp; English.
XX
XX The present sequence is part of a plant RNA-directed RNA polymerase
XX protein. Polynucleotides encoding RNA-directed RNA polymerases were
XX isolated from plant cDNA libraries. They are useful as probes for
XX genetically and physically mapping genes, and as markers for
XX traits linked to those genes. They are useful for controlling gene
XX expression and provide mechanisms to engineer plant virus resistance.
XX They are also useful for plant breeding to develop lines with desired
XX phenotypes.
XX
XX Sequence 417 AA;
XX
XX Query Match 17.3%; Score 1013.5; DB 21; Length 417;
XX Best Local Similarity 48.3%; Pred. No. 6.9e-82;
XX Matches 200; Conservative 80; Mismatches 119; Indels 15; Gaps 8;

QY 498 SRTLSVLRIEIEVDPVKY--HGTSYVPSDGIKTSKSGFAHRVASKGLQ---YTPSAFC 553
Db 7 SRTTFEVSYYDVEIPDIEVTDTGTIYFSDGIGIKTSRFAHQVAKLIGDPAHPPSAFQ 66
QY 554 IRYGGYKGVGVGVPDPSMKLSLRKSKESBNDIKDLVLSWYSKYQPCYLNRLQITLSTL 613
Db 67 IRYGGYKGVITIDPTSEFFNLSLAPSKKFKSKSTMLNITNWSKQPCYVYNREIISLSTL 126
```





```
PS Claim 23; Page 59; 62pp; English.
CC The present sequence is part of a plant RNA-directed RNA polymerase
CC protein. Polynucleotides encoding RNA-directed RNA polymerases were
CC isolated from plant cDNA libraries. They are useful as probes for
CC genetically and physically mapping genes, and as markers for
CC traits linked to those genes. They are useful for controlling gene
CC expression and provide mechanisms to engineer plant virus resistance.
CC They are also useful for plant breeding to develop lines with desired
CC phenotypes.
XX
SQ Sequence 90 AA;

Query Match 3.4%; Score 197; DB 21; Length 90;
Best Local Similarity 48.9%; Pred. No. 1.3e-09;
Matches 44; Conservative 19; Mismatches 23; Indels 4; Gaps 2;

QY 1 MGKTIQVGFPPYLLSAEVKSFLEKTYGTVCALEVKQSKGG-SRAFAKVFADNISAD 59
DQ |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1 MGKTIELGFTSVNSVDKTFVQYTGEGTVFAIKLKHGKGRVPRAFAIIQFTANSAT 60
QY 60 KIITLANN---RLYFGSSYLKAWEMKTDIV 86
DB 61 SMSFRANILTRLRGTSYLKAREMERDIV 90

RESULT 14
AAB28535.
ID AAB28535 standard; Protein; 111 AA.
XX
AC AAB28535;
XX
DT 07-FEB-2001 (first entry)
XX
DE Rice RNA-directed RNA polymerase, SEQ ID NO: 16.
XX
KW Rice; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
KW plant virus resistance; plant breeding.
XX
OS Oryza sativa.
XX
PN WO2000060097-A1.
XX
PD 12-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US09105.
XX
PR 07-APR-1999; 99US-0128094.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI (PION-) PIONEER HI-BRED INT INC.
XX
PI Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;
XX
XX WPI; 2000-679376/66.
DR N-PSDB; AAC63744.
XX
XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
XX for controlling gene expression and providing mechanisms to engineer
XX plant virus resistance.
XX
XX Claim 23; Page 58; 62pp; English.
PS
CC The present sequence is part of a plant RNA-directed RNA polymerase
CC protein. Polynucleotides encoding RNA-directed RNA polymerases were
CC isolated from plant cDNA libraries. They are useful as probes for
CC genetically and physically mapping genes, and as markers for
CC traits linked to those genes. They are useful for controlling gene
CC expression and provide mechanisms to engineer plant virus resistance.
CC They are also useful for plant breeding to develop lines with desired
CC phenotypes.
XX
SQ Sequence 111 AA;

us-09-782-874-2.rag Page 11
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:44:48 : Search time 21 Seconds  
(without alignments)  
1560.816 Million cell updates/sec

Title: US-09-782-874-2

Perfect score: 5864

Sequence: 1 MKGTIQVFPYLLSAEVVK.....RPVLNLSLRAGLSHRLVLK 1114

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA.\*

- 1: /cgn2\_6/prodata2/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/prodata2/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/prodata2/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/prodata2/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/prodata2/iaa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/prodata2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5864	100.0	1114	4	US-08-811-583-2
2	1182	20.2	218	4	US-08-811-583-3
3	113.5	1.9	831	4	US-09-269-861A-8
4	112.5	1.9	988	3	US-08-851-843A-69
5	112.5	1.9	988	4	US-08-874-549A-112
6	112.5	1.9	988	4	US-08-854-050-69
7	112.5	1.9	988	4	US-09-430-323-69
8	109	1.9	1215	4	US-09-134-001C-5319
9	105.5	1.8	659	4	US-09-562-737-11
10	105.5	1.8	1194	2	US-08-488-940-1
11	105	1.8	1181	2	US-08-488-940-2
12	104	1.8	1177	3	US-08-754-490-10
13	104	1.8	1177	3	US-08-922-505A-10
14	104	1.8	1177	4	US-09-260-952A-10
15	104	1.8	1177	4	US-09-253-341-10
16	104	1.8	1177	4	US-09-253-331A-10
17	104	1.8	1177	4	US-09-261-040-10
18	103	1.8	1169	4	US-09-255-829-20
19	103	1.8	1177	3	US-08-754-490-28
20	103	1.8	1177	3	US-08-922-505A-28
21	103	1.8	1177	4	US-09-260-952A-28
22	103	1.8	1177	4	US-09-253-341-28
23	103	1.8	1177	4	US-09-253-331A-28
24	103	1.8	1177	4	US-09-261-040-28
25	102.5	1.7	1194	2	US-08-488-940-17
26	102.5	1.7	1626	2	US-08-771-602B-2
27	101.5	1.7	856	4	US-09-134-001C-5438

28	101.5	1.7	1093	4	US-09-315-793-52	Sequence 52, Appl
29	101.5	1.7	1180	3	US-09-224-024-28	Sequence 28, Appl
30	101.5	1.7	1180	5	PCT-US94-07902-28	Sequence 28, Appl
31	101.5	1.7	1626	4	US-09-232-446B-2	Sequence 2, Appl
32	101	1.7	800	2	US-08-488-940-4	Sequence 4, Appl
33	101	1.7	813	2	US-08-488-940-3	Sequence 3, Appl
34	101	1.7	1071	2	US-08-975-527-1	Sequence 1, Appl
35	100.5	1.7	1194	2	US-08-488-940-18	Sequence 18, Appl
36	100	1.7	1177	3	US-08-754-490-28	Sequence 26, Appl
37	100	1.7	1177	3	US-08-922-505A-26	Sequence 26, Appl
38	100	1.7	1177	4	US-09-260-952A-26	Sequence 26, Appl
39	100	1.7	1177	4	US-09-253-341-26	Sequence 26, Appl
40	100	1.7	1177	4	US-09-253-331A-26	Sequence 26, Appl
41	100	1.7	1177	4	US-09-261-040-26	Sequence 26, Appl
42	100	1.7	1661	2	US-08-882-083-2	Sequence 2, Appl
43	100	1.7	1661	2	US-08-558-107-2	Sequence 2, Appl
44	100	1.7	1661	4	US-09-243-539-2	Sequence 2, Appl
45	99.5	1.7	3169	4	US-09-453-702B-257	Sequence 257, App

ALIGNMENTS

RESULT 1  
US-08-811-583-2  
; Sequence 2, Application US/08811583  
; Patent No. 6218142  
; GENERAL INFORMATION:  
; APPLICANT: Wassenecker, Michael  
; APPLICANT: Riedel, Leonhard  
; APPLICANT: Schiebel, Winfried  
; APPLICANT: Sanger, Heinz  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN  
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RDRP)  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & NEAVE  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/811,583  
; FILING DATE: 05-MAR-1997  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley, James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: MPC-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1114 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-811-583-2

Query Match 100.0%; Score 5864; DB 4; Length 1114;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1114; Conservative 0; Mismatches 0; Gaps 0;  
QY 1 MKGTIQVFPYLLSAEVVKFLEXYGTGYTCALEVKSGSRFAKVFQADNISADK 60  
|||||

Db 1 MGKTIQVFGFPYLLSAEVKSFLEKTYGTGTVCALEVQKSGSRFAKAVQPADNISADK 60  
Qy 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDQMDGTLNFCQISDDKFAVLGSTEVS 120  
Db 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDQMDGTLNFCQISDDKFAVLGSTEVS 120  
Qy 121 IQGIGLKKFFFLSSGADYKQLQSYENIWOVVLHRPYGQNAQFLLIQLFGAPRIYKRL 180  
Db 121 IQGIGLKKFFFLSSGADYKQLQSYENIWOVVLHRPYGQNAQFLLIQLFGAPRIYKRL 180  
Qy 181 ENSCYSFFKTPDDQWVRTTDFPPSWIGLSSSLCLOFRGRCVRLPNPEESFFHYAERENNI 240  
Db 181 ENSCYSFFKTPDDQWVRTTDFPPSWIGLSSSLCLOFRGRCVRLPNPEESFFHYAERENNI 240  
Qy 241 TLQGTFFVYSOKSALPNVQPEGISIPYKILFKISSLVHOGCIPGALNVYFFRLVDP 300  
Db 241 TLQGTFFVYSOKSALPNVQPEGISIPYKILFKISSLVHOGCIPGALNVYFFRLVDP 300  
Qy 301 RRRNVACIEHALEKLYIKKCCYDVPWMLTEQYDGYLKGQPPKSPSITLDDGLVYVRRV 360  
Db 301 RRRNVACIEHALEKLYIKKCCYDVPWMLTEQYDGYLKGQPPKSPSITLDDGLVYVRRV 360  
Qy 361 LVPCKVYFCGPEVNVNRLVNSEIDNLFVRSFVDEWEKLYSTDLLPKASTGSGVR 420  
Db 361 LVPCKVYFCGPEVNVNRLVNSEIDNLFVRSFVDEWEKLYSTDLLPKASTGSGVR 420  
Qy 421 TNYERILSTLRKGFVIGDKKFEFLAFSSQLRDNVSMFASRPGLTANDIRAWMGDFSQ 480  
Db 421 TNYERILSTLRKGFVIGDKKFEFLAFSSQLRDNVSMFASRPGLTANDIRAWMGDFSQ 480  
Qy 481 IKNVAKYAARLGQSGSRETLVLRHEIEVDPVKVHGTSYVFSGIGKISGDFARHVA 540  
Db 481 IKNVAKYAARLGQSGSRETLVLRHEIEVDPVKVHGTSYVFSGIGKISGDFARHVA 540  
Qy 541 SKCGLQYTPSAFQIRYGYGVGVDPDSSMKLSLRKSMKSYSDNKLKDLVGLWSKYQPC 600  
Db 541 SKCGLQYTPSAFQIRYGYGVGVDPDSSMKLSLRKSMKSYSDNKLKDLVGLWSKYQPC 600  
Qy 601 YLNQLITLSTLVQKDEVLEQKQKAVDQDLAILHDSLKAQAELEMSGENTNLIKAM 660  
Db 601 YLNQLITLSTLVQKDEVLEQKQKAVDQDLAILHDSLKAQAELEMSGENTNLIKAM 660  
Qy 661 LNCYKPDABEFLSMQLQTPRASKLLDLRTSRIFIPNGRTMGMCLDESRTLEYGVFVQ 720  
Db 661 LNCYKPDABEFLSMQLQTPRASKLLDLRTSRIFIPNGRTMGMCLDESRTLEYGVFVQ 720  
Qy 721 FTGAGHGFEFSDLLHPFNNSRTSNFNLKGNVYVAKNPCLHPGDIRVLKAVNRALHHV 780  
Db 721 FTGAGHGFEFSDLLHPFNNSRTSNFNLKGNVYVAKNPCLHPGDIRVLKAVNRALHHV 780  
Qy 781 DCVVPQKGRPHNECSGSDLDGDIYFVQWQDMIPPRQVPMYPPAPSIQLDHDTV 840  
Db 781 DCVVPQKGRPHNECSGSDLDGDIYFVQWQDMIPPRQVPMYPPAPSIQLDHDTV 840  
Qy 841 BEVEEYFTNYIVNDSLGIIANAHVVFADREPDMAMSDPCCKLAELFIAVDFFPKTGVPAE 900  
Db 841 BEVEEYFTNYIVNDSLGIIANAHVVFADREPDMAMSDPCCKLAELFIAVDFFPKTGVPAE 900  
Qy 901 IPSQLRPKEYDQWKPDKTSYISERVIGKLFVRKVKQKAPQASSIAFTTRDVARSDAD 960  
Db 901 IPSQLRPKEYDQWKPDKTSYISERVIGKLFVRKVKQKAPQASSIAFTTRDVARSDAD 960  
Qy 961 MEVDGFEDYIDEAFDYKTEYDNKLGNDYIGKTEAELLSSGIMKASKTFDRKDAEAI 1020  
Db 961 MEVDGFEDYIDEAFDYKTEYDNKLGNDYIGKTEAELLSSGIMKASKTFDRKDAEAI 1020  
Qy 1021 SVAVRALKEARAFKRRNDIDMLPKASAWHYHTYHTYWCYNOGLKRAHFTSEFPWCV 1080  
Db 1021 SVAVRALKEARAFKRRNDIDMLPKASAWHYHTYHTYWCYNOGLKRAHFTSEFPWCV 1080  
Qy 1081 YDQLIQIKKDKARNPVLNLSLRAQLSHRLVLK 1114  
Db 1081 YDQLIQIKKDKARNPVLNLSLRAQLSHRLVLK 1114

RESULT 2  
US-08-811-583-3  
; Sequence 3, Application US/08811583  
; Patent No. 6218142  
; GENERAL INFORMATION:  
; APPLICANT: Wassenecker, Michael  
; APPLICANT: Riedel, Leonhard  
; APPLICANT: Schiebel, Winfried  
; APPLICANT: Sanger, Heinz  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMIC ACTIVITY OF AN  
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RDRP)  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & NEAVE  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/811.583  
; FILING DATE: 05-MAR-1997  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley, James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: MPG-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 218 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-811-583-3

Query Match 20.2%; Score 1182; DB 4; Length 218;  
Best Local Similarity 100.0%; Pred. No. 2.3e-112; Indels 0; Gaps 0;  
Matches 218; Conservative 0; Mismatches 0;

QY 700 RTMGMCLDESRTLEYGVFVQFTGAGHGFEFSDLLHPFNNSRTSNFNLKGNVYVAKNP 759  
Db 1 RTMGMCLDESRTLEYGVFVQFTGAGHGFEFSDLLHPFNNSRTSNFNLKGNVYVAKNP 60  
QY 760 LHPGDIRVLKAVNRALHHVDCVVPQKGRPHNECSGSDLDGDIYFVQWQDMIPPR 819  
Db 61 LHPGDIRVLKAVNRALHHVDCVVPQKGRPHNECSGSDLDGDIYFVQWQDMIPPR 120  
QY 820 QVQPMYPPAPSIQLDHDTVIEVEEYFTNYIVNDSLGIIANAHVVFADREPDMAMSDPC 879  
Db 121 QVQPMYPPAPSIQLDHDTVIEVEEYFTNYIVNDSLGIIANAHVVFADREPDMAMSDPC 180  
QY 880 KKLAEFLSTAVDPKTVGVAEIPSQLRPKEYDFDMKXP 917  
Db 181 KKLAEFLSTAVDPKTVGVAEIPSQLRPKEYDFDMKXP 218

RESULT 3  
US-09-269-861A-8  
; Sequence 8, Application US/09269861A  
; Patent No. 6468775  
; GENERAL INFORMATION:



APPLICANT: Ankenbauer, Waltraud  
APPLICANT: Markau, Ursula  
APPLICANT: Svetlichny, Vitaly  
APPLICANT: Schmitz-Ahnehan, Gudrun  
APPLICANT: Reiser, Astrid  
APPLICANT: Angerer, Bernhard  
APPLICANT: Ebenbichler, Christine  
APPLICANT: Laue, Frank  
APPLICANT: Bonch-Osmolovskaya, Elizaveta  
TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASE FROM CARBOXYDOTHERMUS HYDROGENOFORMAN  
FILE REFERENCE: 4494  
CURRENT APPLICATION NUMBER: US/09/269,861A  
CURRENT FILING DATE: 1999-11-22  
PRIOR APPLICATION NUMBER: PCT/EP97/05391  
PRIOR FILING DATE: 1997-10-01  
PRIOR APPLICATION NUMBER: EP/96115873.0  
PRIOR FILING DATE: 1996-10-03  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 8  
LENGTH: 831  
TYPE: PRT  
ORGANISM: Carboxydotherrnus hydrogenoformans  
US-09-269-861A-8

Query Match 1.9%; Score 113.5; DB 4; Length 831;  
Best Local Similarity 19.0%; Pred. No. 0.068;  
Matches 158; Conservative 119; Mismatches 298; Indels 257; Gaps 42;

Qy 410 LPKASTGSGVTNIVERTLSRLKGFVIGDKKFFELFASSQLRDNVWVFASRGLTAN 469  
Db 20 LPPLKTTGEGTGAIVGELTMLFR--VTKDEKPEYLA-----VAFDISKRTETE 67  
Qy 470 DIRAMGDFQIKNKVAKYARLGOSFSGSRETSLVLRHEIVPDVKVHGTSYVFSQIG 529  
Db 68 QPTAYKGRKEAPD-----ELVPQFALVREVVLKN-----VPYELDG--YEADDIIG 114  
Qy 530 KISGDF--HVVASKGLOTPSNFQIRYGGYKGV-----GVDPSSMKLSLRKNSK 581  
Db 115 HLSRAFAGQGEV-----VITADRMQLQVDEKTVVLYLTKGI--TELVKMDLAAILN 167  
Qy 582 Y-----ESDNKIDVLGWSKYQPCYMLNRLITLLSTLGVRKDELEOKQK 625  
Db 168 YGLKPKQLVDYKGLMGDFSDNI-----PGVPGIGKGTALDLIKTVSGVEVLARK-- 217  
Qy 626 EAVDQDLAILHDSLKAQBALEIMSPGENTNLIKAMLC-----GYKPDREPLSM 675  
Db 218 ---DELKPKRLKLAHENLAKISKQLATILREIPLEISLEDLVKKEPNYEEVAKLEFLH 274  
Qy 676 MLOTRASKLLDLRTSRIFIPNGRTMGCLDESRTL--YGVVFQFTGAGHGFSDDLH 734  
Db 275 EFKSP-----LKEIEPKIKKEVQEGKDLV-----QVETVETEGQIAVFSGDFYVDDGETK 326  
Qy 735 PFNNRSSTNSNFIKGNVVVAKNP-----CLHPG-----DIRVLKAV-----N 772  
Db 327 FYSLDRLNEIEIFRNKIIITDARKGIYHVCKLEKLTPEVCFDARIAAYVLNPDQNG 386  
Qy 773 VVALHHWDCVFPQKGRPHNECSGDLGDYF-----VCDQDMIPROVQPMEX 826  
Db 387 LKGLYLKLDLPY-----EDVSLNIRGLFYLKEMMRKPIFEQO--ERLFYEIEL 434  
Qy 827 PPAP-----SIQLDHDVTI-----EEVEEYFTNYVNDLSLGIIANAHVVFADREP 871  
Db 435 PLTPVLAQMEHTGIQVDRALKMSLELGEQIEE-----LIREIYVLAGE--- 479  
Qy 872 DNAMSDPKKLAELSIAVDFPKTGPV-----AEIPSLRP----- 907  
Db 480 EFNLSF--ROLGVIL-----FEKGLPLVIKTKTGYSDAEVLLEBLLPFHEIGIKILNY 533  
Qy 908 -----KEYPDFMDKPKDTSYISERV-----IGKLFKRVKDKAPQASSIATFRD 951  
Db 534 RQLMKLKSYTDLK-----PLINERTKGLHTFTNQGTGLTORLASSPFLQNI--PIRL 586

Qy 952 VARRSYDADMEYDFEDYIDEAFYKTEYDNKLGMLNDYIGIKTEAELISGGIMKASKTF 1011  
Db 587 LGRKLKRMFIPSPGY-DYIVSA-----DYSQIELRLLAHF--SEEPKLEA--YQKEDI 636  
Qy 1012 DRKDAEATSVAVRALRKEARAFKERN-----DIDDM-----LPKASANYHTYHFT 1059  
Db 637 HRKTASEVFGVLEEVTPEMRAHAKSVNFGIVGISDFGLGRDLKIPREVAKYI----- 591  
Qy 1060 YMGYNOGLKRAHFISFPWC--VYDLOIQIKDK-----ARNRPVLNLS 1102  
Db 692 -----KNYFANYPKVREVLDLVTAREKGVYVTLFGRRRYPPELSS 733

## RESULT 4

US-08-851-843A-69  
Sequence 69, Application US/08851843A  
Patent No. 6093809  
GENERAL INFORMATION:  
APPLICANT: Cech, Thomas R.  
APPLICANT: Lingner, Joachim  
APPLICANT: Nakamura, Toru  
APPLICANT: Chapman, Karen B.  
APPLICANT: Morin, Gregg B.  
APPLICANT: Harley, Calvin  
APPLICANT: Andrews, William H.  
TITLE OF INVENTION: No. 6093809el Telomerase  
NUMBER OF SEQUENCES: 225  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/851,843A  
FILING DATE: 06-MAY-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/724,643  
FILING DATE: 01-OCT-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-002930US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 69:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 988 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-851-843A-69

Query Match 1.9%; Score 112.5; DB 3; Length 988;  
Best Local Similarity 18.2%; Pred. No. 0.12;  
Matches 128; Conservative 109; Mismatches 258; Indels 207; Gaps 30;



```
Db 329 DDEKILSYSLKPNQVFAFLRSILV---RVF---PKLIWGNQRIFEIILKOLETEFLKLSRY 382
Qy 398 DE-----EWEKLYSTDLLPKASTGSGVTRNIYERILSTLRKGFVIGDKKF 442
Db 393 ESFSLHYLMSNIKISEIWLVLGRSNKAKMCLSDFERKKQIFAEFYIWLNSFIPILOS 442
Qy 443 EFLAFSSQLRDSNV-----WMFASRPGLTANDIRAW----- 474
Db 443 FFYITESSDLNRRTVYFRKDIWKLCPFITSMKMEAFKINENNVRMDTKTLPPIAVI 502
Qy 475 -----MGDFSQIKNVAK-YAARLQGS-----FGSSRETL-----SVLRH-----EIEVIP-DVK 516
Db 503 RLLPKKNTFRILTNRKFLIKMGSKNKMVSTNQTLRPVASILKHLINESGIPFNLE 562
Qy 517 VHGTSYVSDGIGKISGDFAHRVASKGLQYTPSAFYIRYGGYGVGVDPDSSMKLSLR 576
Db 563 VYMKLLTFKKDLLK-----HRM-----FGKKYFVRIDIKSCY----- 595
Qy 577 KSMKSYESDNKLD-----VLGWSKYQPCYLNRLQITLLSTLGVKDEVLEQKQKEAVDOLD 632
Db 596 -----DRIKQDLMPRIIVKKKLKDPEFVIRKYATIHAT---SDRATKNFVSEAFSYFD 644
Qy 633 AILHDSLKAQEALELMSPGENTNLIKAMLCNGYKPDAPFELSMMLQTFRASKLLDL---R 689
Db 645 MV-----PFEKVYQLLSMKTSDTLFDVFDVYWTKSSSEIF--KMKLHLSGHIVKIGNSQ 697
Qy 690 TRSRIFIPNGRT-----MMGCLDE--SRILEYGQVQVQ 720
Db 698 YLQKVGIPQGSILSSFLCHFYMEDLIDELYSFTKKKGSVLLR 739

RESULT 6
US-08-854-050-69
; Sequence 69, Application US/08854050
; Patent No. 6261836
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. 6261836el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/854,050
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
```

```
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-0029300S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 69:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 988 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-854-050-69
```

```
Query Match 1.9%; Score 112.5; DB 4; Length 988;
Best Local Similarity 18.2%; Pred. No. 0.12;
Matches 128; Conservative 109; Mismatches 258; Indels 207; Gaps 30;

Qy 134 LSSGSADYKQLQSYENITQVVLHPRYQNAQFLLIQLFAGAPRIYKRLNESCYSFFKETPD 193
Db 130 LVSTFPNLYLSILESKNWQ-----LLEIIGSDAMHYLLSKG--SIFPALPN 174
Qy 194 DQWVRTTDFP-----PSWIGLSSSLCLOQFRRGV 221
Db 175 DNYLQISGIPLEKNNVFEETVSKKRTIETSIQNKSAKEVSWNSISISRSFIYRSS 234
Qy 222 RLNFEESEFHYAERENNTIQTGFTFFVQSXSALVNPQPPGISIPYKILEKISLV- 280
Db 235 YKFKQDLVFNLSICDRTNVHMWLOWTFPRQGLNAPQ-----VKLHKVPLVS 286
Qy 281 QHGCIQGPALNVYFFRLVDPRRNRVACIEHALEKLYIKCECCYDPVWMLTEQDGYLKR 340
Db 287 QSTVVPKRLLVY--PLIEQTAKRLHRI--SLSKVY-----NHICYPIDTH 328
Qy 341 QPPKSPSITLDDGLV--YVRVLVTCKYFCGPEVNVSN-RVLRYSEIDINFLRVSVF 397
Db 329 DDEKILSYSLKPNQVFAFLRSILV---RVF---PKLIWGNQRIFEIILKOLETEFLKLSRY 382
Qy 398 DE-----EWEKLYSTDLLPKASTGSGVTRNIYERILSTLRKGFVIGDKKF 442
Db 383 ESFSLHYLMSNIKISEIWLVLGRSNKAKMCLSDFERKKQIFAEFYIWLNSFIPILOS 442
Qy 443 EFLAFSSQLRDSNV-----WMFASRPGLTANDIRAW----- 474
Db 443 FFYITESSDLNRRTVYFRKDIWKLCPFITSMKMEAFKINENNVRMDTKTLPPIAVI 502
Qy 475 -----MGDFSQIKNVAK-YAARLQGS-----FGSSRETL-----SVLRH-----EIEVIP-DVK 516
Db 503 RLLPKKNTFRILTNRKFLIKMGSKNKMVSTNQTLRPVASILKHLINESGIPFNLE 562
Qy 517 VHGTSYVSDGIGKISGDFAHRVASKGLQYTPSAFYIRYGGYGVGVDPDSSMKLSLR 576
Db 563 VYMKLLTFKKDLLK-----HRM-----FGKKYFVRIDIKSCY----- 595
Qy 577 KSMKSYESDNKLD-----VLGWSKYQPCYLNRLQITLLSTLGVKDEVLEQKQKEAVDOLD 632
Db 596 -----DRIKQDLMPRIIVKKKLKDPEFVIRKYATIHAT---SDRATKNFVSEAFSYFD 644
Qy 633 AILHDSLKAQEALELMSPGENTNLIKAMLCNGYKPDAPFELSMMLQTFRASKLLDL---R 689
Db 645 MV-----PFEKVYQLLSMKTSDTLFDVFDVYWTKSSSEIF--KMKLHLSGHIVKIGNSQ 697
Qy 690 TRSRIFIPNGRT-----MMGCLDE--SRILEYGQVQVQ 720
Db 698 YLQKVGIPQGSILSSFLCHFYMEDLIDELYSFTKKKGSVLLR 739
```

RESULT 7  
US-09-430-323-69  
; Sequence 69, Application US/09430323  
; Patent No. 6309667  
; GENERAL INFORMATION:  
; APPLICANT: Cech, Thomas R.  
; Linger, Joachim  
; Nakamura, Toru  
; Chapman, Karen B.  
; Morin, Gregg B.  
; Harley, Calvin  
; Andrews, William H.  
; TITLE OF INVENTION: No. 6309667el Telomerase  
; NUMBER OF SEQUENCES: 225  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/430.323  
; FILING DATE: 29-Oct-1999  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/854,050  
; FILING DATE: 09-MAY-1997  
; APPLICATION NUMBER: US 08/851,843  
; FILING DATE: 06-MAY-1997  
; APPLICATION NUMBER: US 08/846,017  
; FILING DATE: 25-APR-1997  
; APPLICATION NUMBER: US 08/844,419  
; FILING DATE: 18-APR-1997  
; APPLICATION NUMBER: US 08/724,643  
; FILING DATE: 01-OCT-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Apple, Randolph T.  
; REGISTRATION NUMBER: 36,429  
; REFERENCE/DOCKET NUMBER: 015389-002930US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 69:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 988 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 69:  
US-09-430-323-69

Query Match 1.9%; Score 112.5; DB 4; Length 988;  
Best Local Similarity 18.2%; Pred. No. 0.12;  
Matches 128; Conservative 109; Mismatches 258; Indels 207; Gaps 30;  
QY 134 LSSGSADYKQLQSYENINQVHLRHPYQNAQFLIQLFGAPRIYKRLNSCYSFKEPTPD 193  
DB 130 LVSTFPNYLISLESKNQ-----LLEIIGSDAMHYLLSKG--SIFEALPN 174  
QY 194 DQWRTTDFP-----PSWIGLSSSLCLQFRGV 221  
DB 175 DNYLQISGIPFKNNVFEYTSKKKRIETSIQNKSRKEVSNWSISISFIFRYS 234  
QY 222 RLPNPFESFFHYAERENITLQGTFFVQSALVPWQPEGISIPYKILFISSLV- 280  
DB 235 YKKFKODLYFNLSICDNTVMWLQWTFPRQGLINAFQ-----VKQLHKVPLVS 286

QY 281 QHGCIPGALNVYFFRLVDPRRNRVACIEHALEKLYIKECCYDPVRLWTEYDGYLGR 340  
DB 287 QSTVVPKRLKVV--PLIEQTAKRLHRI--SLSKV-----NHYCPYIDTH 328  
QY 341 QPPKSPSITLDDGLV--YVRVLVTPCKYFCGPEVNVSN-RVLNYSSEDIDNEFLRVSEV 397  
DB 329 DDEKILSYSLKPNQVFAFLRSILV---RVF---PKLWGNQRIFEILLKDLTFKLRSY 382  
QY 398 DE-----EWEKLYSTDLLPKASTGSGVTRTIYERILSTLRKGFVIGDKKF 442  
DB 383 ESESLHLYMSNIKISIEIWLVLGKRSNAKCLSDFEKRKQIFAETIYWLNSFIIPILQS 442  
QY 443 EPLAFSSQLRDSV-----WVFASRPGLTANDIPAW----- 474  
DB 443 FFYITESSDLNRRTVYFRKDIWKLCLCRPFITSMKMEAFKINENNVMDTQTTTPPAVI 502  
QY 475 ----MGDFSQIKNVAK-YAARLQGS---FGSSRETL---SVLRH---EIEVIP-DVK 516  
DB 503 RLLPKKNTFRLLTNLRKFLIKMGSNNKMLVSTNQTLRPVASILKHLINEESSGIPFNLE 562  
QY 517 VHGTSYVSDGIGKISGDFAHVASKCGLOYPSPAPQIRYGGYGVGVVDPOSSMKLSLR 576  
DB 563 VYMKLLTFKKDLLK-----HRM-----FGRKKYFVRIDIKSY----- 595  
QY 577 KMSKYESDNILKLD---VLGWSKYQPCVNLRLITLLSTLGVKDEVLEQKQEAVDQLD 632  
DB 596 -----DRIKQDLMFRIVKVKLKDPEVIRKYATIHAT---SDRATKNFYSEAFSYD 644  
QY 633 AILHDSLKAQEALELMSGENTNILKAMLCNGYKPDAPFPLSMLOTFRASKLLDL---R 689  
DB 645 MV-----PFEKVQQLLSMKTSDTLFDVFDYWTKSSEIF--KMLKEHLSGHIVIGNSQ 697  
QY 690 TRSRIPNGRT-----MMGCLDE--SRILEYGVQVFO 720  
DB 698 YLQKGIPOGSISSLFCHFYMEDLIDEYLSFTKKKGSVLLR 739  
RESULT 8  
US-09-134-001C-5319  
; Sequence 5319, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5319  
; LENGTH: 1215  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5319  
Query Match 1.9%; Score 109; DB 4; Length 1215;  
Best Local Similarity 19.1%; Pred. No. 0.4;  
Matches 117; Conservative 85; Mismatches 203; Indels 208; Gaps 31;  
QY 539 VASKCG--LQYTSATQ--TRYGGYKGVGVVDPOSSMKLSLRKSMKYESDNILKLDVLGW 594  
DB 118 IPSRMGLLLDMSPRALEVIFYASY---VVVDPGPT---GLEKKTLLSEAE-----F 163  
QY 595 SKYQPCVNLRLITLLSTLGVKD---EV-LEQKQEAVDQLDAILHDSL-KAQEALELMS 649  
DB 164 REYDYKPN-QFVAKWAGAEIGKLLLEIDLEDELKELRDELESATGRLTRAIKRLVEVE 222  
QY 650 PGENTNILKAMLCNGYKPDAPFPLSMMLQ-----TFRASKLLDLRTR----- 691

Db 223 SFRNGNPNMILDLVLP1PIPTPIRPMVQDGGFATSDLNDLYRRVIRNRRNLKRLDLD 282  
Qy 692 --SRIFIPNGRTMGCIDES--RTLEYGOVFOFTGAGHGEFSDDLHPFNNSRSTNSFI 747  
Db 283 GAPILIVQNEKRM--LOEAVDALIDNGRGRPVTPGNGRPLKSLSHMLKAGKGRFRQNL 339  
Qy 748 L-----KGNVYVAKNP-----CLHPGDIR-----VLK-----AVNVR----- 774  
Db 340 LGRVDYSGRSVIAVGPGLKMYQGLPKEMALELFPVKELVQVRBIATNIKNNAKSIE 399  
Qy 775 -----ALHNM-----VDCVFPQKGRPH--NECSGSDLD 803  
Db 400 RMDDEVWDLVDVTEHPVLLNRAPTHLRIGIQAFETIVEGRAIRLHLPLVTAYNADF 459  
Qy 804 GDIYFV-----CWDQMIPPRQVQPMYPPAPSIQLDHDTVIEEVEEYF 847  
Db 460 GDQMAVHPLSKAQAEARMMLAAQILNPKDKGKPVVTPSQDMVLGNLYTLTLERKDAVN 519  
Qy 848 TNYIVNDS-----LGIANA--HVFEADREPDMAMSDCKKLAEFSI 888  
Db 520 TGAIFNDTNEVLKAYGVHHLTRIGVHANSFNPTTDEQNSKILATSVGKI--IFNE 577  
Qy 889 AVDFPKTGVAEIPSQLRPKEYPDMKPKDKTSYISERVIGKLFKRVKDKAPQASSIATF 948  
Db 578 II-----PDSFAYINEPSQANLERTTPDK----- 601  
Qy 949 TRDVARSYDAD--MEVDGFEDYIDAFDYKTEYDNK-LGNLMDYIGIKTEAELISG-G 1003  
Db 602 -----YFVDPTQLGEGGLKEYFDNA-ELIEFPNKKFLGNII-----AEVFNRS 644  
Qy 1004 IMKASKTFDRKD 1016  
Db 645 ITDISMLDRMKD 657

RESULT 9  
US-09-562-737-11  
; Sequence 11, Application US/09562737  
; Patent No. 6428967  
; GENERAL INFORMATION:  
; APPLICANT: Herz, Joachim  
; TITLE OF INVENTION: IDL Receptor Signaling Pathways  
; FILE REFERENCE: UTSW0708  
; CURRENT APPLICATION NUMBER: US/09/562,737  
; CURRENT FILING DATE: 2000-05-01  
; NUMBER OF SEQ ID NOS: 132  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 11  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: Sequence  
US-09-562-737-11

Query Match 1.8%; Score 105.5; DB 4; Length 659;  
Best Local Similarity 20.5%; Pred. No. 0.29;  
Matches 67; Conservative 46; Mismatches 119; Indels 95; Gaps 16;  
Qy 791 RHPNNECSGSLDGIYFVQWQDMIPPRQVQM-----EYPPAPSIQLDHDTVIE 841  
Db 320 RHP-----SISEDEGDFC-----LSSPERAEPGGGWRGSLGPPPPPPRASLSST 370  
Qy 842 EYEEVFTNYIVNDSIGITANAHV-----FADREPDMAMSDCKKLAEFSIADVFP 893  
Db 371 SYDSVKYTLVDEH-----AQLELSLRPCFGDYSDESATVYNCASASSPYESAI--- 423  
Qy 894 KTGVAEIPSQLRPKEYPDFMDKPKTSYISERVIGKLFKRVKDKAPQASSIATF----- 949  
Db 424 --GEEYEAPQPRPTCLSESDTPDEPDVFSKKEL-NVFMGSRGRSSSAESGFLFSCVIN 480

Qy 950 -----ROVARRSYDADMEVDG-----FEDVIDEAFDYKTEYDNKLGNDY 990  
Db 481 GEEHEQTHAIRFPRHEDELEVEDDPLLVLELQAEADYWEAYNMRT---GARGVFPAY 537  
Qy 991 YGKLT--EAEILSGGIMKASKTFDRKDAEATSVAVRALRKEARAWFKRRNDIDDMPL-- 1046  
Db 538 YALEVTKEPEHMAA-----LAKNSCVLEISVRGVKIGVKA-----DDALEAK 579  
Qy 1047 --KASAWYHVT-----YHP---TYWG 1062  
Db 580 GNKCSHFOLKNISFCGYHPKNNKYFG 606  
RESULT 10  
US-08-488-940-1  
; Sequence 1, Application US/08488940  
; Patent No. 5854049  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Guy L.  
; TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/488,940  
; FILING DATE: 09-JUN-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fraser, Janis K.  
; REGISTRATION NUMBER: 34,819  
; REFERENCE/DOCKET NUMBER: 05433/009001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1194 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-488-940-1

Query Match 1.8%; Score 105.5; DB 2; Length 1194;  
Best Local Similarity 19.9%; Pred. No. 0.89;  
Matches 176; Conservative 114; Mismatches 360; Indels 235; Gaps 43;  
Qy 218 RRGVRLPNFEE-SFFHYAERENNTITQGTFF-----VSQKSALVNPVQPPG-ISP 269  
Db 326 QKGEIMENIPQMSAFWAVRTAVTAVINAASGRQTVDEALKQAQTNSSSVYGRGSIAGP 385  
Qy 270 YKIL---FKISSLVQH--GCIPG--PALNYYFRL-----VDPRRRNA 306  
Db 386 EWLDRSVNNSQLVSVAGTVEGTNQDISLKFFEDLISRPAHGKTEGGLSPKSKPPA 445  
Qy 307 ---CIEHALEKLYIKECCYDPVRLTEQY-----DGYLKGRQPPKSPSITLDDGLV 355  
Db 446 TDSGAMSHKLEKA-----DILLKAIQEQLIANVHNSDDYFEVIDFASDATITDRNGV 497  
Qy 356 YVR-----RVLYTPCKVYFCGPEVNV--SNRVLRYNSEDIDNPLRVSVFDEWEKLYST 407  
Db 498 YFADKDGSVTLPTQPVQGEFLLSGHRVVRVRYKEKPIQNAQKSVSDVEYTVQF-----T 547

```
QY 408 DLLPKASTGSGVRNIYERILSTRKGFVIGDKKFEFLAFSSQLRDNSSVWFASRPGLT 467
Db 548 PLNPDDEFRPLGKDT---KLLKTLAIGDTITSQ--ELLAQAQSILNKN-----HPGYT 595
QY 468 A-----NDIRAWMGDFSQ-----IKNVAKAARLGQSGSRETSLVLRHEIEV 511
Db 596 IYERDSSIVTHDNDIFRILPMDQEFYRVKN-REQAIRINKKSGLNEE-----643
QY 512 IPDVKHGTSYVSGDGIGKISGDFAHVASKGGLQTP-----SAFQIRYGGYKGVGV 565
Db 644 -----INNTDLISE-----KYVVKKEKPYDPDRSHLKLFTIKY-----V 680
QY 566 DPSSMKLSLRKSKSYESNIKLDVLGWSKYQPCYLNRQLITLLSLGVKDEVLEQKQK 625
Db 681 DVDITN-----ELLSQTLTASERNLDFR--DLYDPRDKAKLLYNLDAFGIMDYTLTGKVE 735
QY 626 EAVDQLDAILHDSL-KAQEALELMSPGENTNILKAMLCNGYKPDABPFLSMMLQTFRASK 684
Db 736 DNHDDTNRILTYVMGKRPE-----GENASYHLAYDKDRYTEEREVYSILRYT--GTP 786
QY 685 LLDL---RTSRIFIPNGRTMMGLDESRTLEYQGVFQVGTGAGHGFSDDLHPFNNSRS 741
Db 787 IPDPNPKNSQLVSVAGTVEG-TNQDISLKFEIDLTSRPAHGKTEQGLSPKSKPFA 845
QY 742 TNSNFIKNGVNAKNPCLHPGDIRVLKAV-----NVRALHEMWDVCPVPPQKGRPHN 795
Db 846 TDSG-----AMSHKLEKADL--LKAIQEQLIANVHNSDDYFEVIDF-----AS 886
QY 796 ECGSGLDGIYFCWDDMIPRQVOPMEYPPAPSIQLDHDVYIEVEEYFTNYIVNDS 855
Db 887 DATITDRNGKVFADKDGSVTLPTQ-----PVQEFLLSGHVRV-----RYEKKPIQNOA 935
QY 856 LGLIANAHVAFADREPDMWMSDPCK--KLAELFSIAVDFPKTGVPAPISQLRKPEPDF 913
Db 936 KSVDEVTVQFTPLNPDDFRPLGKDTKLLKTLAIGDTITSQELLAQAQSILN-KNHPGY 994
QY 914 MDXPKDTSYISERYIGLFRKVKDKAPQASIAITFTRDVARRSYDADMEVDGFEDYIDEA 973
Db 995 T-----IYER-----DSSIVTHDNDIFRILPMDQE-----1020
QY 974 FDYKTEYDNKLGMLDYYGIKTE---AEILSGG---IMKASKTFD 1012
Db 1021 FTYRVKNREQAIRINKKSGLNEEINNTDLISEKYIYVLKKEKPYD 1065

RESULT 11
US-08-488-940-2
; Sequence 2, Application US/08488940
; Patent No. 5854049
; GENERAL INFORMATION:
; APPLICANT: Reed, Guy L.
; TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,940
; FILING DATE: 09-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
```

REFERENCE/DOCKET NUMBER: 05433/009001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/542-5070

TELEFAX: 617/542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1181 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-488-940-2

Query Match 1.8%; Score 105; DB 2; Length 1181;

Best Local Similarity 19.9%; Pred.No.0.98; 359; Indels 228; Gaps 42;

Matches 174; Conservative 114; Mismatches 114; PGRSIEGR--N 382

QY 218 RGVRLPNPEE-SFFHYAERENNTITLQG-FTFFVSQKSALVPNVQPPGEGISPIYKILFK 275

Db 326 QKEIMPNIPOMSAFWYAVRTAVINAAAGROFVDEALQDAQTNSSSVPGRSGIEGR--N 382

QY 276 ISSLVQH--GCTPG--PALNVTFFRL-----VDPRRNVA---CIEHAL 312

Db 383 NSQLVSVAGTVEGTNQDISLKFEIDLTSRPAHGKTEQGLSPKSKPAFADSGAMSHKL 442

QY 313 EKLYIKKECCYDPVRLWLTQY-----DGLYKGRQPKPSITLDDGLVYVR-----R 359

Db 443 EKA-----DLLKAIQEQLIANVHNSDDYFEVIDFASDAITIDRNGKVIYFADKDGSVT 494

QY 360 VLVTPCKVYFCGPEVNV--SNRVLRNYSIEDINFLRVSPVDEWEKLYSTDLLPKASTGS 417

Db 495 LPTQPVQEFLLSGHVRVRYKPKIQNAKSDVVEYTVQF-----TLPNPDODFRP 544

QY 418 GYRTNIYERILSTRKGFVIGDKKFEFLAFSSQLRDNSSVWFASRPGLTA-----468

Db 545 GLKDT---KLLKTLAIGDTITSQ--ELLAQAQSILNKN-----HPGYTIYERDSSIYV 592

QY 469 --NDIRAWMGDFSQ-----IKNVAKAARLGQSGSRETSLVLRHEIEVDPVKVHGS 521

Db 593 HNDIFRITLPMDDQEFYRVKN-REQAIRINKKSGLNEE-----INNTD 635

QY 522 YVSDGIGKISGDFAHVASKGGLQTP-----SAFQIRYGGYKGVGVDPDSSMKLSL 575

Db 636 LISE-----KYVVKKEKPYDPDRSHLKLFTIKY-----VDVDTN--EL 674

QY 576 RKSMKSYESDNTKLDVLGWSKYQPCYLNRQLITLLSLGVKDEVLEQKQEAVDQLDAIL 635

Db 675 LKSEQLLTASERNLDFR--DLYDPRDKAKLLYNLDAFGIMDYTLTGKVEDHDDTNRIL 732

QY 636 HDLSL-KAQEALELMSPGENTNILKAMLCNGYKPDABPFLSMMLQTFRASKLLDL---RTR 691

Db 733 TVYMGKRPE-----GENASTHAYDKDRYTEEREVYSILRYT--GTFIPDNPNKNN 783

QY 692 SRIFIPNGRTMMGLDESRTLEYQGVFQVGTGAGHGFSDDLHPFNNSRSTNSFILKN 751

Db 784 SOLVSVAGTVEG-TNQDISLKFEIDLTSRPAHGKTEQGLSPKSKPAFADSG-----836

QY 752 VVYAKNPCLHPGDIRVLKAV-----NVRALHEMWDVCPVPPQKGRPHNEGSGSLDGD 805

Db 837 ---AMSHKLEKADL--LKAIQEQLIANVHNSDDYFEVIDF-----ASDAITIDRNGK 883

QY 806 IYFVCWDDQMIPRQVQPMYPPAPSIQLDHDVYIEVEEYFTNYIVNDSLGIIANAHVY 865

Db 884 VYFADKDGSVTLPTQ-----PVQEFLLSGHVRV---RYEKKPIQNAKSDVVEYTVQ 932

QY 866 FADREFDMAMSDPK--KLAELFSIAVDFPKTGVPAPISQLRKPEYPPFMDKPKDTSYI 923

Db 933 FTPLNPDDDFRPLGKDTKLLKTLAIGDTITSQELLAQAQSILN-KNHPGYT-----I 983

QY 924 SERVIGKLFKRVKDKAPQASIAITFTRDVARRSYDADMEVDGFEDYIDEAFDYKTEYDNK 983

Db 984 YER-----DSSIVTHDNDIFRILPMDQE-----FTYRVKNREQ 1017



## TELECOMMUNICATION INFORMATION:

TELEPHONE: (512)418-3000  
TELEFAX: (512)474-7577  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1177 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-922-505A-10

Query Match 1.8%; Score 104; DB 3; Length 1177;  
Best Local Similarity 18.2%; Pred. No. 1.2;  
Matches 168; Conservative 112; Mismatches 301; Indels 340; Gaps 45;

QY 102 FGCQISDDKFAVLGSTEVS-----IQFGIGLKKEFFFLSSGSDYKQLQSYENIMQVVLH 156  
DB 13 YNC-LSNPEVVLGGERIETGYTIDISLTQFLSEFVPGAGFVLGL-VDIINGI--- 67

QY 157 RPYG--QNAQFL-----IQLGAPRIYKRL--NSCYSPFKETPDQWVRTTDFP 203  
DB 68 --FGPSQWDAFLVQIQLNORIEEFARNOAISRLGLSNLIQIYAES-----PREWEAD 120

QY 204 PSWIGLSSLCLOPRGVRLENPEESFFHYAERENNITLQGTFFVFSQKSALVPNVQPP 263  
DB 121 PTNPALREEMRIQF-----NDMNSALTTAIPLFA-----VQNYQVP 156

QY 264 -----EGISIPYKILFKISSLVQHGCIPGALN----- 291

DB 157 LLSVVYQAAHLHLSVLDRVSVFGQWGFDAATINSRYNDLTRLIGNYDYAVRWYNTGLE 216

QY 292 -----VYFRLVDPRRNVACIEHALEKLYIKKECCYD 324

DB 217 RVWGPDSDRWVRYNQFRRELTFLVLDIVALFPNYDSRRYPIRTVSQLTREIY-----TN 270

QY 325 PVRWLTQEDGYLKG-----RQPKSP-----SITL-----DDGLVY--VRRVLVTPCK 366

DB 271 PV---LENFDGSPRGSAGQIERSIRSHPLMDILNSIIITDHRGYYIWSGHQIMASP-- 325

QY 367 VYFCGPEVNVSNRVLNYSIEDINFLRVSVFDEWEKLYST--DLLPK-----ASTGSGVR 420

DB 326 VFGSGPEFTF----- 358

QY 421 TNIVERILSTL-RKGFVIGDKKEFFLAFSSQLR--DNSVWVFASRPGLTA----- 468

DB 359 ---YRTLSTLYRRPFNIG-----INNQLSLVDGTEFAFGTSSNLPASVYRKSGTV 407

QY 469 -----NDIRAWMGDFSOIKNVAKYAARLGQSGSRETLVSLR----- 506

DB 408 DSLDEIPPPQNNVPPRQGFSHRLSHVSMFRS-----GFSNSSVSIIRAPMFSWTHRSAT 461

QY 507 -----HEIEVIPDKVH-----GTSYVPSDGI--GKI-----SGFAHRVASKCGLQYT 548

DB 462 PTNTIDPERITQIPLVKAHTLQSGTTVVRGPGFTGGDILRRTSGGPFATIVNINGQLPQ 521

QY 549 PSFAQIRYGG-----YKGVGV-----DPDSSMKLSLRKSKSY----- 582

DB 522 YRARIYASTNLRIYTVAGERIFAGQFNKMTDGLPTFFQSFYSATINTAFTEPMSQ 581

QY 583 ESDNIKLDVLGWSKQPCYLN-----OLITLLSTLGVKDE 618

DB 582 SSFTVGADT--FSSGNEYIIDRFELIPVTAITEAYDLERAQKAVNALFTSINQIGIKTD 639

QY 619 VLEQKQKAEVDOLDAI-----LHDSLKAQBAELMSFGENTNLKAKLNCGYKPDPAEPLSM 675

DB 640 VTDYH-----IDQVSNLVDCLSDEFCLDEKRELSEKVKHAKRLSDERNLLQDPNFKGINRQ 695

QY 676 MLQFTFRASKLLDTRSRIFTPNRGTMMGCLDES-RILEYQCV-----FVQFTGAGH 727

DB 696 LDRGWKSGTDITIQGGDVFKEKYVTLPGTDECYPTLYQKIDESKIKAFTRYQLRGYI 755

QY 728 EFSDDLHPFNNSRSTNSNFIKLVNNAKNCPLHPGDIRVLKAVNRALHHMVDVCFVFPQ 787

DB 756 EDSQDL-----EIVLIRYN---AKH-----ETVNVPGTGLWPLSAQSP 791

QY 788 KGRPHNECS-----GSDLD 803

DB 792 IGKCGEPNRCAPHLEWNPDL 812

RESULT 14  
US-09-260-952A-10  
Sequence 10, Application US/09260952A  
Patent No. 6221649  
GENERAL INFORMATION:  
APPLICANT: Malvar, Thomas  
APPLICANT: Gilmer, Amy Jelen  
TITLE OF INVENTION: HYBRID BACILLUS THURINGIENSIS DELTA-ENDOTOXINS WITH  
TITLE OF INVENTION: NOVEL BROAD SPECTRUM INSECTICIDAL ACTIVITY  
FILE REFERENCE: MECO:217  
CURRENT APPLICATION NUMBER: US/09/260,952A  
CURRENT FILING DATE: 1999-03-02  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 10  
LENGTH: 1177  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: coding  
US-09-260-952A-10

Query Match 1.8%; Score 104; DB 4; Length 1177;  
Best Local Similarity 18.2%; Pred. No. 1.2;  
Matches 168; Conservative 112; Mismatches 301; Indels 340; Gaps 45;

QY 102 FGCQISDDKFAVLGSTEVS-----IQFGIGLKKEFFFLSSGSDYKQLQSYENIMQVVLH 156  
DB 13 YNC-LSNPEVVLGGERIETGYTIDISLTQFLSEFVPGAGFVLGL-VDIINGI--- 67

QY 157 RPYG--QNAQFL-----IQLGAPRIYKRL--NSCYSPFKETPDQWVRTTDFP 203  
DB 68 --FGPSQWDAFLVQIQLNORIEEFARNOAISRLGLSNLIQIYAES-----PREWEAD 120

QY 204 PSWIGLSSLCLOPRGVRLENPEESFFHYAERENNITLQGTFFVFSQKSALVPNVQPP 263  
DB 121 PTNPALREEMRIQF-----NDMNSALTTAIPLFA-----VQNYQVP 156

QY 264 -----EGISIPYKILFKISSLVQHGCIPGALN----- 291

DB 157 LLSVVYQAAHLHLSVLDRVSVFGQWGFDAATINSRYNDLTRLIGNYDYAVRWYNTGLE 216

QY 292 -----VYFRLVDPRRNVACIEHALEKLYIKKECCYD 324

DB 217 RVWGPDSDRWVRYNQFRRELTFLVLDIVALFPNYDSRRYPIRTVSQLTREIY-----TN 270

QY 325 PVRWLTQEDGYLKG-----RQPKSP-----SITL-----DDGLVY--VRRVLVTPCK 366

DB 271 PV---LENFDGSPRGSAGQIERSIRSHPLMDILNSIIITDHRGYYIWSGHQIMASP-- 325

QY 367 VYFCGPEVNVSNRVLNYSIEDINFLRVSVFDEWEKLYST--DLLPK-----ASTGSGVR 420

DB 326 VFGSGPEFTF----- 358

QY 421 TNIVERILSTL-RKGFVIGDKKEFFLAFSSQLR--DNSVWVFASRPGLTA----- 468

DB 359 ---YRTLSTLYRRPFNIG-----INNQLSLVDGTEFAFGTSSNLPASVYRKSGTV 407

QY 469 -----NDIRAWMGDFSOIKNVAKYAARLGQSGSRETLVSLR----- 506

DB 408 DSLDEIPPPQNNVPPRQGFSHRLSHVSMFRS-----GFSNSSVSIIRAPMFSWTHRSAT 461

QY 507 -----HEIEVIPDKVH-----GTSYVPSDGI--GKI-----SGFAHRVASKCGLQYT 548

DB 462 PTNTIDPERITQIPLVKAHTLQSGTTVVRGPGFTGGDILRRTSGGPFATIVNINGQLPQ 521



QY 549 PSFQIRYGG-----YKGVVGV-----DPSSMKLSLRKMSKY----- 582  
Db 522 RYRARIYASTNRIYVTVAGERIFAGQFNKMTDGTPLTFQSFYSATINTAFTFPMQ 581  
QY 583 ESDNIKLDVLGWSKYQPCYLNR-----OLITLLSTLGVKDE 618  
Db 582 SSFTVGADT--FSSGNEVYIDRFELIPVTATFEAYDLERAQKAVNALFTSINGIKTD 639  
QY 619 VLEQKQKEAVDQDAI---LHDSLKAQEALELMSPGENTNLIKAMLCGYPDAEPFLSM 675  
Db 640 VTDYH-----IDQVSNLVDCLSDEFCLDEKRELSKVHAKRLSDERNLLQDPNFKGINRQ 695  
QY 676 MLOTFRASKLLDLRTSRIFIPNGRTMMGCLES--RTLEYGOV-----FVQFTGAGHG 727  
Db 696 LDRGWRGSTDTIQRGDDVFKENYVTLPGTDECYPYLYQKIDESKLKAFTRQLRGYI 755  
QY 728 EFSDDLHPFNNSRSTNSNFIKGNVVAKNPCLPHPGDIRVLKAVNRALHHMDCVFPQ 787  
Db 756 EDSQDL-----EYILIRYN---AKH-----ETVNVPGTGSMLPLSAQSP 791  
QY 788 KGKRPHPNECS-----GSDLD 803  
Db 792 IGKCGEPNRCAPHLEWNPDL 812

RESULT 15

US-09-253-341-10  
; Sequence 10, Application US/09253341  
; Patent No. 6242241

GENERAL INFORMATION:

APPLICANT: Maivaer, Thomas  
; Gilmer, Amy Jelen  
TITLE OF INVENTION: BROAD-SPECTRUM (-ENDOTOXINS  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/253,341

FILING DATE: 10-Feb-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/922,505

FILING DATE: 03-SEP-1997

ATTORNEY/AGENT INFORMATION:

NAME: Kitchell, Barbara S.

REGISTRATION NUMBER: 33,928

REFERENCE/DOCKET NUMBER: MECO:163

TELECOMMUNICATION INFORMATION:

TELEPHONE: (512)418-3000

TELEFAX: (512)474-7577

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1177 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 10:

US-09-253-341-10

Query Match

Best Local Similarity 18.2%; Score 104; DB 4; Length 1177;

Matches 168; Conservative 112; Mismatches 112; Indels 340; Gaps 45;

Search completed: November 6, 2002, 03:47:17

Job time : 28 secs

QY 102 FGQIISDDKFAVLGSTEVS-----IQFGIGLKKFFFLSSGSADYKQLQLSYENIQWVLH 156  
Db 13 YNC--LSNPEVEVLGGERIETGYTPIDISLTOFLLSEFVPGAGFVLGL--VDIWI-- 67  
QY 157 RPYG--QNAQFL-------TOLFGAPRIYKRL--NSCYFFKRETPDDQVVRTDPP 203  
Db 68 --FGPSQWDAFLVQIEQLINORIEEFARNQAIISRLGLESLNLYQIYAES-----FREWEAD 120  
QY 204 PSWIGLSSSLCLOFRRGVRLPNFEESFFHYAERENITITQGTFTFFVSOKSALVPPNQPP 263  
Db 121 PTNPALAEEMRIQF-----NDMSALTATPLFA-----VQNVQVP 156  
QY 264 -----EGISIPYKILFIKISSLVQHGCIPGALN----- 291  
Db 157 LLSVYVQAANLHSLVRDVGFGQRMGDAATINSRYNDLTRIGNYTDYAVRWYNTGLE 216  
QY 292 -----VYFFVLVDPRRRNVACIEHALEKLYIYKECCYD 324  
Db 217 RVNGPDSRWVRVNRQFRELTLVLVDIVALFPNYSRRYPPIRVISQLTRIY-----TN 270  
QY 325 PVRWLITEQDGYLKG-----RQPKSP-----SITL-----DDGLVY--VRRVLVTPCK 366  
Db 271 PV---LENFDGSPRGSQAQIERSIRSPHLMIDILNSITITDAHRGYYVWSGHQIMASP-- 325  
QY 367 VYFCGPEVNVSNRVLNRYSEDIDNFLRVSVFDEWEKLYST--DLLPK-----ASTGSGVR 420  
Db 326 VGFSGPPEFTF-----PLYGTMGNAAPQOQRIIVAQLOGGV- 358  
QY 421 TNYIERILSTL-RKGFVIGDKKFEFLAFSSQLR--DNSVYMFASRPLFA----- 468  
Db 359 --YRTLSSSTLYRRFPNIG-----INNOQLSVLDGTEFAYGTSNLPSSAVYRKSGTV 407  
QY 469 -----NDIRAWMGDFSQIKNAVAKYAAARLQSGSSRETLISVL- 506  
Db 408 DSLDEIPQNNVPPRQGFSHRLSVSMFRS-----GFSNSVSVIIRAPMFWSTHRSAT 461  
QY 507 -----HEIEVIPDVKVH-----GTSYVFSQGI--GKI-----SGDFAHRVASCKGLQYT 548  
Db 462 PTNIDIPERTQIDPLVKAHLQSGTIVVVGFTGGDILRRTSGGPFAYITVINGOLPQ 521  
QY 549 PSFQIRYGG-----YKGVVGV-----DPSSMKLSLRKMSKY----- 582  
Db 522 RYRARIYASTNRIYVTVAGERIFAGQFNKMTDGTPLTFQSFYSATINTAFTFPMQ 581  
QY 583 ESDNIKLDVLGWSKYQPCYLNR-----OLITLLSTLGVKDE 618  
Db 582 SSFTVGADT--FSSGNEVYIDRFELIPVTATFEAYDLERAQKAVNALFTSINGIKTD 639  
QY 619 VLEQKQKEAVDQDAI---LHDSLKAQEALELMSPGENTNLIKAMLCGYPDAEPFLSM 675  
Db 640 VTDYH-----IDQVSNLVDCLSDEFCLDEKRELSKVHAKRLSDERNLLQDPNFKGINRQ 695  
QY 676 MLOTFRASKLLDLRTSRIFIPNGRTMMGCLES--RTLEYGOV-----FVQFTGAGHG 727  
Db 696 LDRGWRGSTDTIQRGDDVFKENYVTLPGTDECYPYLYQKIDESKLKAFTRQLRGYI 755  
QY 728 EFSDDLHPFNNSRSTNSNFIKGNVVAKNPCLPHPGDIRVLKAVNRALHHMDCVFPQ 787  
Db 756 EDSQDL-----EYILIRYN---AKH-----ETVNVPGTGSMLPLSAQSP 791  
QY 788 KGKRPHPNECS-----GSDLD 803  
Db 792 IGKCGEPNRCAPHLEWNPDL 812

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:45:23 ; Search time 13 Seconds  
(without alignments)  
1188.490 Million cell updates/sec

Title: US-09-782-874-2

Perfect score: 5864

Sequence: 1 MGKTIQVGPYLLSAEVK.....RPVLNLSRAQLSHRLVLK 1114

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 90412 seqs, 13869272 residues

Total number of hits satisfying chosen parameters: 90412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5864	100.0	1114	10	US-09-782-874-2
2	1182	20.2	218	10	US-09-782-874-3
3	112.5	1.9	1184	10	US-09-815-242-5229
4	112.5	1.9	1188	10	US-09-815-242-12125
5	109.5	1.9	1579	10	US-09-801-368-368
6	107	1.8	1357	10	US-09-815-242-11997
7	105.5	1.8	1080	10	US-09-781-558-4
8	105.5	1.8	1135	10	US-09-737-149-35
9	104.5	1.8	711	10	US-09-966-561-2
10	104	1.8	1177	10	US-09-873-873-10
11	104	1.8	1177	10	US-09-873-873-12
12	104	1.8	1177	10	US-09-873-873-14
13	103	1.8	1177	10	US-09-873-873-28
14	102	1.7	897	10	US-09-815-242-11609
15	102	1.7	1331	10	US-09-801-368-370
16	101.5	1.7	1115	10	US-09-781-558-2
17	100	1.7	1177	10	US-09-873-873-26
18	99.5	1.7	914	10	US-09-823-356-8
19	99.5	1.7	914	10	US-09-922-217-1066

20	99.5	1.7	914	10	US-09-833-263-1066
21	99.5	1.7	2813	10	US-09-381-261A-1
22	99	1.7	660	12	US-10-115-178-1
23	98.5	1.7	932	10	US-09-788-657-17
24	98.5	1.7	932	10	US-09-788-657-18
25	97	1.7	1019	10	US-09-434-066-23
26	96.5	1.6	624	10	US-09-815-242-5640
27	96.5	1.6	625	10	US-09-815-242-12704
28	96	1.6	1502	10	US-09-801-368-140
29	95.5	1.6	736	10	US-09-815-242-11293
30	95.5	1.6	892	10	US-09-203-658-42
31	95.5	1.6	892	10	US-09-844-353A-42
32	95.5	1.6	991	10	US-09-815-242-5803
33	95.5	1.6	5701	10	US-09-864-761-37319
34	95	1.6	17	10	US-09-782-874-12
35	95	1.6	595	10	US-09-925-297-604
36	95	1.6	869	10	US-09-815-242-10623
37	95	1.6	1216	10	US-09-815-242-13312
38	94.5	1.6	552	10	US-09-817-742-4
39	94.5	1.6	709	10	US-09-815-242-13349
40	94	1.6	935	10	US-09-757-781-1
41	94	1.6	1186	10	US-09-826-752-8
42	93.5	1.6	601	10	US-09-815-242-12358
43	93.5	1.6	1206	12	US-10-032-717-2
44	93.5	1.6	1216	10	US-09-815-242-13698
45	93.5	1.6	2813	10	US-09-886-900-2

#### ALIGNMENTS

#### RESULT 1

US-09-782-874-2

; Sequence 2, Application US/09782874

; Patent No. US20010023067A1

; GENERAL INFORMATION:

; APPLICANT: Wassenegeger, Michael

; Schiebel, Winfried

; Sanger, Heinz

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING

; POLYPEPTIDES HAVING THE ENZYMIC ACTIVITY OF AN

; RNA-DIRECTED RNA POLYMERASE (RRP)

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESS: RISH & NEAVE

; STREET: 1251 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10020

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/782,874

; FILING DATE: 08-Feb-2001

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/811,583

; FILING DATE: 05-MAR-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Haley, James F.

; REGISTRATION NUMBER: 27,794

; REFERENCE/DOCKET NUMBER: MPG-1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-596-9000

; TELEFAX: 212-596-9090

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1114 amino acids

Query Match 100.0%; Score 5864; DB 10; Length 1114;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKTIQVFGFPYLLSAEVAWSFLEKTYGYGTVCALVCALEVKQSGSRAFAKQVQFADNISADK 60  
DB 1 MGKTIQVFGFPYLLSAEVAWSFLEKTYGYGTVCALVCALEVKQSGSRAFAKQVQFADNISADK 60

QY 61 IITLANNRLYFGSSYLKAWENKTDIVQLRAYVDQMDGTTLNFQCOISDDKFAVLGSTEVS 120  
DB 61 IITLANNRLYFGSSYLKAWENKTDIVQLRAYVDQMDGTTLNFQCOISDDKFAVLGSTEVS 120

QY 121 IOFGIGLKKFFFLSSGSADYKLOLSYENTWQVYLHRYPGONAQFLLIQLGAPRIYKRL 180  
DB 121 IOFGIGLKKFFFLSSGSADYKLOLSYENTWQVYLHRYPGONAQFLLIQLGAPRIYKRL 180

QY 181 ENSCYSFEKTPDQWVRTTDFPSSWIGLSSSLCQFRRGVRLNFEESEFFHYAERENNI 240  
DB 181 ENSCYSFEKTPDQWVRTTDFPSSWIGLSSSLCQFRRGVRLNFEESEFFHYAERENNI 240

QY 241 TLQGTFFVSKSALVPNVOPPEGISIPYKILFKISLSVOHGCIPOPALNVYFFRLVDP 300  
DB 241 TLQGTFFVSKSALVPNVOPPEGISIPYKILFKISLSVOHGCIPOPALNVYFFRLVDP 300

QY 301 RRRNVACIEHALEKLYIKECCYDPVRLWTEQYDGYLKGROPKPSITLDDGLVYVRRV 360  
DB 301 RRRNVACIEHALEKLYIKECCYDPVRLWTEQYDGYLKGROPKPSITLDDGLVYVRRV 360

QY 361 LVTCKVYFCGPEVNVSNRVLNRYSEDIDNPLRVSFVDEWEKLYSTDLLPKASTGSGVR 420  
DB 361 LVTCKVYFCGPEVNVSNRVLNRYSEDIDNPLRVSFVDEWEKLYSTDLLPKASTGSGVR 420

QY 421 TNYIERILSTLRKGFVIGDKKFEFLAFSSQLRDNVWVFASRGLFANDIRAWMGDFSQ 480  
DB 421 TNYIERILSTLRKGFVIGDKKFEFLAFSSQLRDNVWVFASRGLFANDIRAWMGDFSQ 480

QY 481 IKNVAKYAARLQSGFSRRETLVLRHIEIYPVKVHGTSYVPSDGGIGKISGDFAHRA 540  
DB 481 IKNVAKYAARLQSGFSRRETLVLRHIEIYPVKVHGTSYVPSDGGIGKISGDFAHRA 540

QY 541 SKCGLQYTPSAFQIRYGGYKGVGVDPSMKSLSRKSMSKYESDNKLDVLGWSKYQPC 600  
DB 541 SKCGLQYTPSAFQIRYGGYKGVGVDPSMKSLSRKSMSKYESDNKLDVLGWSKYQPC 600

QY 601 YLNROLITLLSTLGVKDEVLPQKEAVDQDLDAIHLDSLKAQEALELMSPGENTNILKAM 660  
DB 601 YLNROLITLLSTLGVKDEVLPQKEAVDQDLDAIHLDSLKAQEALELMSPGENTNILKAM 660

QY 661 LNCGYKPAEPPFLSMMLQTFRASKLLDLTRSRIFIPNGRTMGLDESRTLEYGVQFVQ 720  
DB 661 LNCGYKPAEPPFLSMMLQTFRASKLLDLTRSRIFIPNGRTMGLDESRTLEYGVQFVQ 720

QY 721 FTGAGHGFEFDDLHPFNNSRSTNSFILKGNVYVAKNPCLHPGDIRVLKAVNVRALHHV 780  
DB 721 FTGAGHGFEFDDLHPFNNSRSTNSFILKGNVYVAKNPCLHPGDIRVLKAVNVRALHHV 780

QY 781 DCWVFPQKGRPHNPNCSGSLDGDYFVCWDQMDIPPRQVQPMYPPAPSIQDHDVTI 840  
DB 781 DCWVFPQKGRPHNPNCSGSLDGDYFVCWDQMDIPPRQVQPMYPPAPSIQDHDVTI 840

QY 841 ESEVEEFTNYIVNDSIGLIIANAVVFAADREPDMASDPCKKLAELFSTAVDFPKTGPAP 900  
DB 841 ESEVEEFTNYIVNDSIGLIIANAVVFAADREPDMASDPCKKLAELFSTAVDFPKTGPAP 900

QY 901 IPSQLRPKEYPDMKPKDTSYISERVIGKLFVRKVKAPQASSIATFTREDVARRSYDAD 960  
DB 901 IPSQLRPKEYPDMKPKDTSYISERVIGKLFVRKVKAPQASSIATFTREDVARRSYDAD 960

QY 961 MEVDGFEDYIDAFDYKTEYDNKLGNDYGYGKTEAEILSGIMKASKTFDRRKDAEAI 1020  
DB 961 MEVDGFEDYIDAFDYKTEYDNKLGNDYGYGKTEAEILSGIMKASKTFDRRKDAEAI 1020

QY 1021 SVAVRALRKEARAWFKRRNDIDDMPLKASAWHYVTHPTYWGYNQGLKRAHFISFPWCV 1080  
DB 1021 SVAVRALRKEARAWFKRRNDIDDMPLKASAWHYVTHPTYWGYNQGLKRAHFISFPWCV 1080

QY 1081 YDOLIOIKKDKARNRPVNLSSLRALSHRLVLK 1114  
DB 1081 YDOLIOIKKDKARNRPVNLSSLRALSHRLVLK 1114

RESULT 2  
US-09-782-874-3  
: Sequence 3, Application US/09782874  
: Patent No. US20010023067A1  
: GENERAL INFORMATION:  
: APPLICANT: Wassenegger, Michael  
: Riedel, Leonhard  
: Schiebel, Winfried  
: Sanger, Heinz  
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN  
: RNA-DIRECTED RNA POLYMERASE (RdRP)  
: NUMBER OF SEQUENCES: 13  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: FISH & NEAVE  
: STREET: 1251 Avenue of the Americas  
: CITY: New York  
: STATE: New York  
: COUNTRY: USA  
: ZIP: 10020  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patent In Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA: US/09/782,874  
: APPLICATION NUMBER: US/09/782,874  
: FILING DATE: 08-Feb-2001  
: CLASSIFICATION: <Unknown>  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/811,583  
: FILING DATE: 05-MAR-1997  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Haley, James F.  
: REGISTRATION NUMBER: 27,794  
: REFERENCE/DOCKET NUMBER: MPG-1  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 212-596-9000  
: TELEFAX: 212-596-9090  
: INFORMATION FOR SEQ ID NO: 3:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 218 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: <Unknown>  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-782-874-3

Query Match 20.2%; Score 1182; DB 10; Length 218;  
Best Local Similarity 100.0%; Pred. No. 5.1e-100;  
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 RTMMGCLDESRTLEYGVQFVQFTGAGHGFEFDDLHPFNNSRSTNSFILKGNVYVAKNP 759  
DB 1 RTMMGCLDESRTLEYGVQFVQFTGAGHGFEFDDLHPFNNSRSTNSFILKGNVYVAKNP 60

QY 760 LHPGDIRVLKAVNVRALHHVDCWVFPQKGRPHNPNCSGSLDGDYFVCWDQMDIPPR 819  
DB 760 LHPGDIRVLKAVNVRALHHVDCWVFPQKGRPHNPNCSGSLDGDYFVCWDQMDIPPR 819

```
Db 61 LHPGDIRVLKAVNRALHHMVDVVFQKGRKRPNECSGSDLDGDIYFVCWDQDMIPPR 120
QY 820 QVQPMYPPAPSIOLDHVDVTEEVETNYVNDLSGLITANAHVVFADREPDMAMSDPC 879
Db 121 QVQPMYPPAPSIOLDHVDVTEEVETNYVNDLSGLITANAHVVFADREPDMAMSDPC 180
QY 880 KKLAEFLSIADFFPKTGPAPISQLRPKKEYPDMFMDKP 917
Db 181 KKLAEFLSIADFFPKTGPAPISQLRPKKEYPDMFMDKP 218

RESULT 3
US-09-815-242-5229
; Sequence 5229, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5229
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5229

Query Match 1.9%; Score 112.5; DB 10; Length 1184;
Best Local Similarity 18.2%; Pred No. 0.28;
Matches 162; Conservative 124; Mismatches 299; Indels 305; Gaps 40;

QY 324 DPVRM-LTEQDGYLGRQPPKSPSITLDGLVYVRRVLVTPCKYFCGPEVNSNRVL 382
Db 42 DAIKWVLGEQSAKSLGSK-----MED-----IIFSGAE-----HRKAQ 75
QY 383 NYSE---DIDNPLRVSVFDEWEKLYSDLLPKASTGSGVRTNIVERILSLRKGFVIGD 439
Db 76 NYAEVQLRLDNHSHKLSVDE-----NEVIVT-RLDYSRSGE 109
QY 440 KKEFLAFSSQLRNSVNMFAASRPLGTANDITRMWGDFFSQTKN-----V 484
Db 110 SEY-VINNDRAKUKADIADFLDSGLGKEAYSIIIS-QGRVDEILNAPKIDRRQIIESAGV 167
QY 485 AKYARLGQSGSSRETLSVLREHTEVTPDVKVHGTSTVFSDDGIGKISGDFAHVRVASKG 544
Db 168 LKYYKKAESLNKLQDTEDLNTRVEDIILYDLE-----GRVEPLKEEAIAIAK-- 213
QY 545 LQYTSFAFQIRYGGYGVGVVDPDSSMKLSRKSMSKYESDNKIDVLGWSKYQPCYLN 604
```

```
Db 214 -EYKTLISHQMKHSDI--VVTV-----HDIDQYTNDRQLD----- 245
QY 605 QLITLLSTLGVKDEVLEQKQKEAVDO-----LDAILHDSLKAQEALELMSPGEN 653
Db 246 ORLNDLQGOQANKEADKQRLSQOIQYKGRHOLDNDVESLNYQLYKATEFEKYI--GQ 303
QY 654 TNIL-----KAMLNGYKPDAPFPLSM-----LQTFRASKLLDLTRSRIFIP 697
Db 304 LNVLEERKKNQSETNARYEEQENLIELLENISNEISEAQDTYKSLKSKQKELNAVY-- 360
QY 698 NGRMTMGCLDESRTLEVGVQVFTGAGHGFSDDLHPFNNSRSTNSNFTLKGNVVVAKN 757
Db 361 -----RELE-EQDIY-----SDEAHDEKLEIKNEYITLMSQSDVNN 397
QY 758 POLHPGDIRVLK-----AVNRALHHMVDVVFQ-----KGRKRPNECS 798
Db 398 -----DIRFLKHTIENEAKSKRLDSRLVE--VFQKLDIQOQIKTKKEYQQTNKELS 449
QY 799 GSD-----LDGDIYFVCWDQDMIPPRQVQPMYPPAPSIOLDHVDVTEEVETNYI-- 851
Db 450 AVDKREIKNIEKDLDTKKAQNEYEEKLYQAYRTEKMKTRIDSLATQEEETVFFNGVKH 509
QY 852 ---VNDLSGLITANAHVVFADREPDMAMSDPKKLAELFSIADFFPKTGPAPISQLRP 907
Db 510 ILKAKNKELKGIHGAVAELIDVPSKLTQATETALGASLQHVIVDSEKQGRQA----- 561
QY 908 KEYPDFMDKPKDTSYISERVIGK-LPRKYKDKAPQASSIATFTEDVARRSYDADMEVDGF 966
Db 562 -----IQFLKERNLGRATELPL--NVIOQRVAVDAIDIKSIK-----EANGF 600
QY 967 EDYIDEAFDYKTEYDNKLGML-----DYY-----GIKTEAEI----- 999
Db 601 ISTASEAVKVAPEYQNIIGNLGNITIVDLKHANELARAIRKVRTRIVTLEGDIVNPGGS 660
QY 1000 LSGIMKASTFDRRKDAEAIISAVRALRKEARAKWRKRNDDIDMLPKASAWHVYHPT 1059
Db 661 MTGGARKSKSILSKD-ELTTM-----RHQLEDYLQRTES----- 695
QY 1060 YMGYNQGLKRAHFISPPWCYDQLIQIKKDKARNRPVNLNLSLRAQLSH 1109
Db 696 ----FEQPFKELKKS-----DQLSLEYFEKSKKH-----NLTKEQVHH 730

RESULT 4
US-09-815-242-12125
; Sequence 12125, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
```

; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12125  
; LENGTH: 1188  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12125

Query Match 1.9%; Score 112.5; DB 10; Length 1188;  
Best Local Similarity 19.2%; Pred. No. 0.84;  
Matches 162; Conservative 124; Mismatches 299; Indels 305; Gaps 40;  
QY 324 DPVRW-LFEQDGLVKGQPKSPSITLDDGLVTVRRVLVTPCKVFCGPEVNVNRVLR 382  
DB 43 DAIKWVLGEQSAKSLRGSK-----MED-----IIFSGAE---HRKAQ 76  
QY 383 NYSE---DIDNPLRVSVFDEWEKLYSTDLLPKASTGSGVETNIYERILSTLRKGFVIGD 439  
DB 77 NYAEVQLRLDNRHKKLSYDE-----NEVIVT-RRLYRSGE 110  
QY 440 KKFEPLAFSSQLRNDSVWVFASRPLGTANDIRAWMGDFSOIKN-----V 484  
DB 111 SEY-YINDRARLKDIALELDLSDGLGRAYSIIIS-QGRVDEILNAKPIDRQIIIEGAGV 168  
QY 485 AKYAARLQCSGSRSETLSVLRHIEVIPDVVKVGTSTVYFSDGKGKSGDPAHRVASKCG 544  
DB 169 LKYKRKAESLKNLQDTEEDTRVEDILYDLE-----GRVEPLKEEAIAIAK-- 214  
QY 545 LQYTPSAFQIRYGGYGVGVDPSSMKLSLRKSMKSYESDNKIDVLGWSKYQPCYLNR 604  
DB 215 -EYKTLHQMKHSDI--VVTV-----HDIQYTDNRQLD----- 245  
QY 605 QLITLLSTLGVKDEVLQKQKAVDQ-----LDAILHSLKAQEAELMSPEEN 653  
DB 247 QRLNDLQOQANKEADKORLSQIQYKGRHQLDNVDSLYQVLKATEAFERYT--GQ 304  
QY 654 TNIL-----KAMLNCGYKPAEPFLSMW-----LQTFRASKLLDLRTSRIFIP 697  
DB 305 LNVLEERKNQSEINARVEEQENLIELLENISNEISQAQTYLSKSKQKELNAVI--- 361  
QY 698 NGRTWGMGLDSRTLEYQGVFVQTGAGHGFSDDLHPFNNSRTSNFILKGNVVAKN 757  
DB 362 -----RELE-EQLV-----SDEAHKELEIKNEYITLMSQSDVNN 398  
QY 758 PCLHPGDIIRVLK-----AVNVALHHMVDVCFPQ-----KGRPHPECS 798  
DB 399 -----DIRLUKHTTEENAEKSLRDSRLVE--VFPEQKDOIQOIKTKKEYQOTNKELS 450  
QY 799 GSD-----LDGDIYFVCDQMIPRQVQPMYPPAPSIQLDHDVTIEVEVEYFTNYI-- 851  
DB 451 AVDKKIKNIKLDLTKKAQNEYEEKLYQAVRYTEKMKTRIDSLATQEEVEYTFNGVKH 510  
QY 852 ----VNDLSLGIANAHVAFBREDPMANSDCKKLAEFLSTAVDFPKTGPVPAETPSQLRP 907  
DB 511 ILKANKKELKGHGAVAEIIDVPKSLQAIETALGASIQHVIVDSEKDGQA----- 562  
QY 908 KEYPDFNDPKQKTSYISERVIK-LFRKVKDKAPQASSIATFTRDVARRSYDADMEVDGF 966  
DB 563 -----IQPLKERNLGRATFLPL--NVIQSRVAVTDIKSIK-----EANGF 601  
QY 967 EYIDAEADYKTEYDNKLGNUM-----DYY-----GKTEAEI----- 999  
DB 602 ISIASEAVKVAPEYQNIIGNLGNITIIYDHLKHANELARAIRYTRIVTLEGDIVNPGGS 661  
QY 1000 LSGGTMKASKYFDRKDAEATISVAVRALRKEARAFKRRNDIDMDLPKASAWYHVYHPT 1059  
DB 662 MFGCGARKSKSLSKD-ELTWM-----RHQLEDYLRGTES----- 696  
QY 1060 YWCGVNOGLKRAHFTSFPCWYVDQLIQKDKARNRPNVLNSSLRAQLSH 1109  
DB 697 ----FEQGFKELKIKS-----DQLSLEYFEKSKH-----NLTKEQVHH 731

RESULT 5  
US-09-801-368-368  
; Sequence 368, Application US/09801368  
; Patent No. US20020128250A1  
; GENERAL INFORMATION:  
; APPLICANT: Busby, Robert  
; APPLICANT: Cali, Brian  
; APPLICANT: Hecht, Peter  
; APPLICANT: Holtzman, Doug  
; APPLICANT: Madden, Kevin  
; APPLICANT: Maxon, Mary  
; APPLICANT: Milne, Todd  
; APPLICANT: Royer, John  
; APPLICANT: Salama, Sofie  
; APPLICANT: Sherman, Amir  
; APPLICANT: Silva, Jeff  
; APPLICANT: Summers, Eric  
; FILE REFERENCE: 109272.147  
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
; CURRENT APPLICATION NUMBER: US/09/801,368  
; CURRENT FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 09/487,558  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: US 60/160,587  
; PRIOR FILING DATE: 1999-10-20  
; NUMBER OF SEQ ID NOS: 440  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 368  
; LENGTH: 1579  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-09-801-368-368

Query Match 1.9%; Score 109.5; DB 10; Length 1579;  
Best Local Similarity 19.2%; Pred. No. 0.84;  
Matches 247; Conservative 184; Mismatches 447; Indels 409; Gaps 73;

QY 15 SAEVWSFLEKYCYGTVCALVYKQSGSRAPAKVQFADNIS---ADKTIITLA---NNR 68  
DB 474 AAEIKSLVDKY-----YRVNLNLPNKRMAERKPKITKTEAFRRNR 513  
QY 69 LYFGSSYLK-AWEMKTDIVOLRAVY-----DQMDGITLN-----FGCQ 105  
DB 514 IDTLNKLWLNKFNEDINIAVLKKWIVGNKELESTTEVDNTVNLDDPAVFATNCKRFAEQ 573  
QY 106 ISDDKFAVLGSTEVSQFGIG---LKKFFFLSSGADYKLQLSY-ENWQVVLHHPYQG 161  
DB 574 IMKEKDIEL-IFOKKIFFPLAPWILKAKFFFLKYQKWTNELNLSYLDQDLEFLLMFPMRL 632  
QY 162 NAQFLLIQLFGAPRIYKRLNESCYSFFKETPPDQWVRTDPPSPWIGLS-----SSLC 214  
DB 633 VKDILILRLSYA---KKIQNPILMKIDQMDDF-----STYIKLAVQMKFTVASYC 680  
QY 215 LQFRGVRVL-PNEE-----ESFFHYAERENNITLOTGFTFFVSOKSALVPWQPEGI 266  
DB 681 NDWFFKVIDPEDHTVVEGLEYYFFSILE-----LRILYSKNSFKTSKEP--DL 728  
QY 267 SIYPKILF-----KISSLYOHGCIPOGALNLYVFFRL--VDPRRRV 305  
DB 729 LLKYWFRNVGYIIDAGELIAAEFTKLTLRVH-----RLHAYLLRQONTPKLENE 782  
QY 306 ACIEHALEKUYI-----KECCYDVPVWLTQEYD----- 334  
DB 783 AAAAEKVLQVFEILGSMKRLNFTNLTAKFQNFVRYKIEDHNYLLKQKETHFLIYT 842  
QY 335 -GYLK--GRQPPKSPSI--TLDDGLVTVRRVLVTPCKVFCGPEVNVNRVLRNYSIEDID 389  
DB 843 GGLEQNGTYLIGSPPELLGCKDDDIIRIKNSDIGDLV---PKLEINN-----SLTIY 893  
QY 390 NFLRVSVDEWEKLYSTDLLPKASTGSGYR---TNIY---ERILSTLRKGFVIGDKKFEF 444

```
Db 894 NAL-----DDWNNS-----NSLSDSISNDGTPPYIKNDITTOPRSY-NGNRVNR 939
Qy 445 LAFSSQRLDNRVWFMFASRPG-----LTANDIRAWMGDFSQIKNVAKYAARLQSGFS 497
Db 940 PDFNSRSTEEFELETRLSGLVVLITPQEPILWEGEMYNL-----S 984
Qy 498 SREILSVLRHIEVIP-----DVKVIGTSVFS-----DGKIGKISG-----DFAHRVAS-----KCG 544
Db 985 DNKTIKPEGLNLKIPNSIDLKCGSSYALSYQCDRFQOISGSSVSLEKSSSETYKNN 1044
Qy 545 LOYTPSA-FOIRYG--GYKGW--GVDRDSSMKLS-----LRSKMSKVES 584
Db 1045 LORINKAYFCTYSVLKNYTKIVTFKVSVDNLLNIFLFGDRDGLNFRINVANNEK 1104
Qy 585 DNKLDV-----LWISKY--OPCYLNRLI-----TLSTGLGVKDEV 619
Db 1105 RSIILLMRLSIGWLKPLAEDCDPTQORVERWCVTSMFAHMHVSGWNILALDECQFSS 1164
Qy 620 LEOKOKEAVDOLDALHDSLRKQEALELMSGENTNILKMLNCGYKPDAPPL--SMML 677
Db 1165 LKQISECMSSL--ISHFDIIGARSIEV-----EKIN-QQARSNL-----DLEDVDDMMML 1213
Qy 678 QT---FRASKLDRTRSRIFPNGRMTMGCLDESRTLEYQGVFQVQFTGAGHGFSDDLH 734
Db 1214 QVNSEFRVQSIWELEERIK--RNPHQTGKVIDDS--DKGNKYLVSLASSISNVS--- 1263
Qy 735 PENSRSNTSNFI---LGNVWVAKNCPCLHPCDIRVLKAVNVRALHHMVDVFPQKGR 791
Db 1264 ---NRWOKRNFPGGTFGRVYSVD--LDNGEILAVKEINIQDSKNQK--IFPLIKEE 1315
Qy 792 ---PHNECS--GSDLDGDIYFVCWDQDMIPPROVQPMYPPAPPSIO--LDHD-VT 839
Db 1316 MSVLILNHPNIVSYGVEVHRDKVNF-----MEYCEGSLAALLEHGRIE 1362
Qy 840 IEVEEYFTNTVNDLSGLIIANAHVVFADREPDMAMSDPCKKLAEIIVDFPKTGVA 899
Db 1363 DEMVQVTLQLL-EGLAYHESGIVHRDKPE-----NILDF--NGV-- 1403
Qy 900 EIPSQLRPEKYPDFMDKPKDTSYSERSVIGLFRKVKDKAPQASSIATFTTRDVARRSYDA 959
Db 1404 -----IKYVDF-----GAKKIANNNGTFLASMKIE----- 1429
Qy 960 DMEVDGFDYIDEAPDYKTEYDNKLNLMYIG--IKTEAILSGGINKASKTDRKDA 1017
Db 1430 --NADGEHEDVTHVSDSRKAVKNE--NALLDMGTPMYMAPESITGSTTKGLGAD--DV 1483
Qy 1018 EAVISAVRALRKEARAFKRRNDIDMLPKASANVHTVHYTHYTGCVYNOGL----- 1068
Db 1484 WSLGCVVLEMITGTRPWANLNEALNTHVAAG--HTPQFPTKDEVSSAGMKFLERCLIQ 1541
Qy 1069 ---KRAHFISF---PWCYVDLIQTK 1089
Db 1542 NPSKRASAVELLMDPW-----IVQIRE 1563
```

## RESULT 6

```
US-09-815-242-11997
; Sequence 11997, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
```

```
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11997
; LENGTH: 1357
; TYPE: PRN
; ORGANISM: Pseudomonas aeruginosa
; US-09-815-242-11997
```

```
Query Match 1.8%; Score 107; DB 10; Length 1357;
Best Local Similarity 18.9%; Pred. No. 1.1;
Matches 156; Conservative 107; Mismatches 268; Indels 296; Gaps 40;
```

```
Qy 331 EYD-----GYLK-----GROPKSPSTLDDGLVYVRRVLVTPCKYVFCGPVNVSNRVL 381
Db 398 ERYDLSAVGKMKNNRRIGTEIEGPGVLSKEDIIDVLKTL-----VDIRNG-- 443
Qy 382 RNYSEIDINP--LRVSFVDEWEKLYSTDLKPKASTGSGVRYTNIYRILSTLRKGFVIG 439
Db 444 KGIVDDIDHGLNRRVRCVGEMAENQFRVGLV-----RVERAVKERL----- 484
Qy 440 KKEFLAFSSQLRDNDSVWFASFAPGLTANDIRAWMGDFSQIKNVAKYAARLQSGSSR 499
Db 485 -----SMAESEGLMPQD-----LINAKPVAIAKEFFGSSQ 515
Qy 500 -----ETLSVLRIEIVIPDKVHGTSYFSDGIGKISGDFAHRVASKCGIQYTPSA 551
Db 516 LSQFMDQNNFLSEITHKRRV-----SALGPGGLTRERAG 549
Qy 552 FQIR-----YGGYKGVGVVDPSMKLSRKMSKYESNKLKDLVGLWSKYQPCYLNQ 605
Db 550 FEVRDVHPTHYG--RVCPITETPEGPNIGLSIATYARTN-KYGFLE-SPYR--VVKDS 602
Qy 606 LIT-----LLSTLGVKDEVLEQ-----KQEAVDOLDAILH-----DSLKAQEALELM--SP 650
Db 603 LVTDIEIVFLSAIEADHIVIAQASATLNEKQLVDELVAVRHLNEFVYKAPEDVTLMDVSP 662
Qy 651 GENTNILKMLNCGYKPDAPFLS-----MMLOTFRASKLL-----DL 688
Db 663 QVVYVVAASLI-----PFLEHDDANRALMGSNMQROAVPTLRADKPLVGTGMERNV 713
Qy 689 RTRSRIFPNGR-TMGCCLDESRTLEYGVFQVQFTGAGHGFEESDDLHPFNNSRSTNSN-- 745
Db 714 ARDSGVCVWARRGVIDSDVASRVV-----VRVADDEVETGEAGVDI--YNLTKYTSRQN 767
Qy 746 -----FILKGNVW-----VAKNCPCLHPGDIRVLKAVNVRALHHMVDVYVFFQKGRPH 793
Db 768 TCINQRLVSKGDVARGDILADGSPSTMDEL-----ALGQNMVAFMPWNGENFE 818
Qy 794 PNECSGDLGDGIYFVCWDQDMIPPROVQPMYPPAPPSIQLDHVDVTEEEVEEYFTNIVN 853
Db 819 DS-----ICLSERVVOEDRFTTHIQELTVCVARDTKLGEETIADIPN-VGE 864
Qy 854 DSLGIIANAHVTFADREPDMAMSDPCKKLAEIIVDFPKTGVPAPISQLRPFKEYPDF 913
Db 865 AALNKLEAGIYVGAEVQ-----AGDILVGKVTPKGEQLTPEE----- 904
Qy 914 MDKPKDTSYSERSVIGLFRKVKD---KAPQAS-----STATTRDVARRSYDA---DM 961
```

```
Db 905 -----KLLRAIFGEKASVDKDTSLRVPTGKGVIVDQVETRGVDSRALSIEKM 956
QY 962 EVDGFEDYIDEADYKTEYDNKLGNDYYGKTEAEIISGIMKASKTFDRRKDA--EA 1019
Db 957 QLDQIRKDLNEEF-----RIVEGA-----IFERLRAALVGA 987
QY 1020 ISVAVRALRK-----BARAWFKR-----NDIDDMPLKASAW 1051
Db 988 KAEGGPALKRGTEITDYLGLRGQWFKLRMADDAALNEQLEKAQAY 1034

RESULT 7
US-09-781-558-4
; Sequence 4, Application US/09781558
; Patent No. US20020034778A1
; GENERAL INFORMATION:
; APPLICANT: Beasley et al.
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000756
; CURRENT APPLICATION NUMBER: US/09/781,558
; CURRENT FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: US 60/199,811
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: US 09/641,426
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1080
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-781-558-4

Query Match 1.8%; Score 105.5; DB 10; Length 1080;
Best Local Similarity 18.5%; Pred. No. 1;
Matches 135; Conservative 102; Mismatches 249; Indels 245; Gaps 37;

QY 55 NISADKIITLANNRLYFGSSYLKAWEMKTDIVOLRAYVD-QMDGI-----TLNFGCOIS 107
Db 249 NIT-DFLLLTENNKFHLESVINITANLSSTKDLLSLFQVMDNIRNSTTVMWFGC--- 304
QY 108 DDKFAVLGSTEVSIOFGIGLKKFFFTSSGSADYKLO-----LSYENIWQVVLHRPY 159
Db 305 -DMSIRQIFEMSTQGLSPPELHWLGDSQNVBELRTEGLPLGLIAHGKTTQSVFEY-Y 362
QY 160 GQNAQFLLIQLFG-APRIYKLENSCYSPFKETPDQDQWRTDPP-----PSWI 207
Db 363 VQDAMELVARAVATATMIQPEL-----ALLPSTMNCMDVKITNLTSGQYLSRFLANTTFR 417
QY 208 GLSSSLCLOFRGRVLRPLNPFESFFHYAERENNI--TLQGTFTFFVSQKSALVPNVQPEG 265
Db 418 GLSGSIKV---KGSTIIISSENNFTWNLQHPMGKPMWTRLGSMQGGRIVMDSGINPEQA 474
QY 266 -----ISIPKILFKISSLVQH-----GCIPGPAALVYFFELVDPRRNVACI 308
Db 475 QRKHTFHQPNKHLRVVTLIEHPFVFTREVDDGLCPAGOL-----CLDPMTNDSMML 528
QY 309 EHALEKLY-----YIKECCY-----DPVRWLTE-----QYDGLKGRQPKPSITLDD 352
Db 529 DELFSSLSHSSNDYPIKFKKCCYCYCIDLLEQLAEDMNFDFDLIVG-----D 576
QY 353 G-----LVVVRVRLVTPCKVY-----FCGPEVNVSNRVL-----RNY 384
Db 577 GRYGAWKNHGTWGLVGDLLSGTANMAVTSFSINTARSOVIDFTSPFTSTSLGILVRTRDT 636
QY 385 SEDIDNEL-----RVSFVDEWEKLYSDLPKASTGSGVRTNIYER 426
Db 637 AAPICAFWPLHWTWMLGIFVALHITAFILFLYEKWS-----PFGMTPKGRNRN--- 685
QY 427 ILSLTKRGFVIGDKKFEFLAFSSQLRDNYSYMFASRPLTA---NDIRAMWGDFSQTKN- 483
```

```
Db 686 -----KVFSF-----SSALNVCYALLF-----GRTAALKPKCWTGRF--LMNL 722
QY 484 -----VAKYAARL-----GOSFGSSRRTLS----- 503
Db 723 WAIFCMFCLSTVTANLAAMVMEKIYEELSGIHDPKLHHPQSGFRFGTVRESSAEDYVRQ 782
QY 504 -----VLRHEIEVIPDKVHGTSYVFSOG-----IGKISGDFAHRYASKGGLQYT 548
Db 783 SPENMEHYNRRYNVPATPD-----GVQYLNKDEPKLDARIMDKALLDYEVSDADCKLLTV 838
QY 549 PSAFOIR-YGGYKGVGVDPDSSMKLSLRKSKSYESNIKLDYL--GWSKYQPCYLNQRQ 605
Db 839 GKPFALIEGYG-----IGLPPNSPLTSNISELSQYKSHGF-MDVLRDKWKYVVPCKGRSF 892
QY 605 LITLLSTLGVK 616
Db 893 AVTETLQMGIK 903

RESULT 8
US-09-737-149-35
; Sequence 35, Application US/09737149
; Patent No. US20020077486A1
; GENERAL INFORMATION:
; APPLICANT: Spaderna, Steven K
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Muralidhara, Padigaru
; APPLICANT: Spytek, Kimberly A.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-620 CIP
; CURRENT APPLICATION NUMBER: US/09/737,149
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/170,564
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: 60/173,165
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,362
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,544
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/174,404
; PRIOR FILING DATE: 2000-01-04
; PRIOR APPLICATION NUMBER: 60/174,962
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 60/223,929
; PRIOR FILING DATE: 2000-08-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1135
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-737-149-35

Query Match 1.8%; Score 105.5; DB 10; Length 1135;
Best Local Similarity 18.5%; Pred. No. 1;
Matches 135; Conservative 102; Mismatches 249; Indels 245; Gaps 37;

QY 55 NISADKIITLANNRLYFGSSYLKAWEMKTDIVOLRAYVD-QMDGI-----TLNFGCOIS 107
Db 275 NIT-DFLLLTENNKFHLESVINITANLSSTKDLLSLFQVMDNIRNSTTVMWFGC--- 330
QY 108 DDKFAVLGSTEVSIOFGIGLKKFFFTSSGSADYKLO-----LSYENIWQVVLHRPY 159
Db 331 -DMSIRQIFEMSTQGLSPPELHWLGDSQNVBELRTEGLPLGLIAHGKTTQSVFEY-Y 388
QY 160 GQNAQFLLIQLFG-APRIYKLENSCYSPFKETPDQDQWRTDPP-----PSWI 207
Db 389 VQDAMELVARAVATATMIQPEL-----ALLPSTMNCMDVKITNLTSGQYLSRFLANTTFR 443
QY 208 GLSSSLCLOFRGRVLRPLNPFESFFHYAERENNI--TLQGTFTFFVSQKSALVPNVQPEG 265
```



```
Db 444 GLSGSIK---KGSTIISSENNFFIWNLOHDPMGKPMWIRLGSWQGRIVMDSGIWMPEQA 500
QY 266 -----ISIPKILFKISSLVQH-----GCIPGALNVYFRLVDPRRNVACI 308
Db 501 QRKTHQHPNKLRLVTLIEHPFVTRVDDGLCPAGOL-----CLDPMTNDSMML 554
QY 309 EHALEKLY-----YKECCY-----DPVRWLTE-----QYDGLKGRQPPKSPSITLDD 352
Db 555 DRFLSSLHSSNDTVPKFKKCCYCIDLLEQLAEDNFDLDLYVG-----D 602
QY 353 G-----LVYVRVLVTPCKY-----FCGPEVNVSNRVL-----RNY 384
Db 603 GYGAWXNGHWGLVGLLSTANNAVTSINTARSQVIDFSPFESTSLGILVTRDT 662
QY 385 SEDIDNEL-----RVSFVDEWEKLYSTDLLPKASTGSGVRTNIYER 426
Db 663 AAPIGAFMWPLHWMLGIFVALHITAFILTYEWS-----PFGMTFKGGRN-----711
QY 427 ILSTLRKGFVIGDKKFEFLAFSSQLRDNVWMPASRGLTA--NDIRAWMGDFSQIKN- 483
Db 712 -----KVFSF-----SSALNVCYALLF-----GRATAKPKCWTGRF--LMNL 748
QY 484 -----VAKYAARL-----GOSFGSRRTEL-----503
Db 749 WAIFCMCLSYTANTLAAMVGEKIYEELSGIHDPKLLHPSQGFRTGVRESSAEDYVRQ 808
QY 504 -----VLNRETEVPDVKHGTSYVFSDG-----IGKISGDPFAHRVASKCGLQYT 548
Db 809 SFPEMHEMYRRYNNVPATPD-----GVQYKNDPEKLDAFIMDKALLDYEVSIDACKLLTV 864
QY 549 PSAFOIR-YGYKGVGVDPDSSMKLSLRKSMKSYESDNIKLDVL--GWSKYOPCYLNQ 605
Db 865 GKPAIEGYC-----IGLPPNSPLTNSISELSQYKSHGP-MDVLHDKWKYVVPCKGRSF 918
QY 606 LITLLSTLGVK 616
Db 919 AVTETLQMGK 929

RESULT 9
US-09-966-561-2
; Sequence 2, Application US/09966561
; Patent No. US2002090696A1
; GENERAL INFORMATION:
; APPLICANT: University of Southern California
; APPLICANT: Miller, Carol A.
; APPLICANT: Dong, Zhao Hui
; TITLE OF INVENTION: APOPTOSIS INHIBITION
; FILE REFERENCE: 13761-724
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/09/966,561
; PRIOR FILING DATE: EARLIER APPLICATION NUMBER: 09/419,694
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-966-561-2

Query Match 1.8%; Score 104.5; DB 10; Length 711;
Best Local Similarity 20.2%; Pred. No. 0.62;
Matches 53; Conservative 40; Mismatches 99; Indels 71; Gaps 13;

QY 781 DCVVPKQKRPHPNECSGDLGDGVFCVWDQDMTPPRVQPM-----EYPPAS 831
Db 314 DPAAIPSTAGRPHPS---ISEEEG---FDC-----LSPERAEPGGGWRGSLGEPpppr 364
QY 832 IQLDHDTVIEVEEYFTNYIVNDSLGIIANAHVY-----FADREPDMAMSDPCKKIA 883
Db 365 ASLSDTSALSYDSVKVTLVDEH-----AQLELVSLRCPCEGSDSDSNATVYDNCASVS 420
```

```
QY 884 ELFSIAV-----DFPKTGYPAEIPSLRKEYPDFDMKDPKDTYSISERVIGLFRKVKDK 938
Db 421 SPYESAIGEYBEAPRPQPPACLSEDSTP-----DEPD--VHFSKKFL-NVFMGRGR 470
QY 939 APOASSIAFT-----RDVARRSYDADMEVDG-----FEDYIDEAFD 975
Db 471 SSSAESFGLFSCIINGEEQEQTHRAIFRVPRHDELELEVDPLLVE-CAEDYWEAYN 530
QY 976 YKTEYDNKLGNDYYGIKTEAE 998
Db 531 MRT---GARGVPAYAYAEVTK 550

RESULT 10
US-09-873-873-10
; Sequence 10, Application US/09873873
; Patent No. US20020064865A1
; GENERAL INFORMATION:
; APPLICANT: Malvar, Thomas
; APPLICANT: Gilmer, Amy Jelen
; TITLE OF INVENTION: Polynucleotide Compositions Encoding Broad-Spectrum S-Endotoxin
; FILE REFERENCE: MECO-210--2
; CURRENT APPLICATION NUMBER: US/09/873,873
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: US 09/253,341
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: US 08/922,505
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: US 08/754,490
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hybrid Delta-Endotoxin
US-09-873-873-10

Query Match 1.8%; Score 104; DB 10; Length 1177;
Best Local Similarity 18.2%; Pred. No. 1.6;
Matches 168; Conservative 112; Mismatches 301; Indels 340; Gaps 45;

QY 102 FCCQISDDKFAVLGSTEVS-----IQFGIGLKKFFFFSSGADYKQLQLSYENIMQVILH 156
Db 13 YNC-LSNPEVEVLGGRIETGTPIDISLQFLLSLSEFVPGAGFVLGL-VDIIMG 67
QY 157 RPYG--QNAQFL-----IQLFGAPRIYKRL--NSCYSEFFKETPDDQWRTTDP 203
Db 68 --FGPSQWDAFLVQIEQLINQRIEFARNQASIRLEGLSNLYQIYAES-----FREWAD 120
QY 204 PSWIGLSSSLCLOFRGVRLPNFEESFFHYABRENNITLQGTFTFFVSQKSALVPNVQPP 263
Db 121 PTNPALREEMRIQF-----NDMNSALTTAIPLFA-----VQNYQVP 156
QY 264 -----EGISIPYKILFKISSLVQHCCIPGALN-----291
Db 157 LLSVVYQAANLHSLVLRDVSFVGQKRGDATTINSRYNDLTRIGNYTDYAVRWNTGLE 216
QY 292 -----VYFFRLVDPRRNVACIEHALEKLYIIECCYD 324
Db 217 RVWGPDSDRWRYNQFRRELTTLVDLVALFPNYSRRYPRTVSQLTREIY-----TN 270
QY 325 PYRWLTEQYDGLKG-----RQPKSP-----SITL-----DDGLVY--VVRVLVTPCK 366
Db 271 PV--LENFDSGFRGSAQGIERSIRPHLMIDILNSITITDAHRGYIYWSGHQIMASP-- 325
QY 367 VYFCGPEVNVSNRNLNYSIEDIDNLRVSVFVDEWEKLYST--DLLPK-----ASTGSGVR 420
Db 326 VGFSGPETTF-----PLYGTMGNAAPQORIVAQLGQV- 358
```



; PRIOR APPLICATION NUMBER: US 08/922,505  
; PRIOR FILING DATE: 1997-09-03  
; PRIOR APPLICATION NUMBER: US 08/754,490  
; PRIOR FILING DATE: 1996-11-20  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 14  
; LENGTH: 1177  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Hybrid Delta-Endotoxin  
US-09-873-873-14

Query Match 1.8%; Score 104; DB 10; Length 1177;  
Best Local Similarity 18.2%; Pred. No. 1.6;  
Matches 168; Conservative 112; Mismatches 301; Indels 340; Gaps 45;

QY 102 FCGQISDDKFAVLGSTEVS-----IQFGIGLKKFFFLSSGSADYKQLQLSYENIMQVVLH 156  
Db 13 YNC-LSNPEVEVLGGRIETGYTPIDISLSLTQFLSEFVPGAGFVLGL-VDIIWGI--- 67  
QY 157 RPYG--QNAQFLL-----IQFGAPRIYKRL--NSCYSEFKETPDQWVRTTDFP 203  
Db 58 --FGSQWDAFLVQIEQLINQRIEFAFNQAIISRLGSLNYQIYAES-----FREWEAD 120  
QY 204 PSWIGLSSSLCLOFRGVRPLNPFEEFHYAERENNTITQGTFFFSQKSAVPNVQPP 263  
Db 121 PTNPALREEMRIQF-----NDMNSALTTPALFA-----VQNVQVP 156  
QY 264 -----EGISIPYKILFKISSLVQHGCCIPGALN----- 291  
Db 157 LLSVYVQAANLHLSVLDRVSVFGQWGFDAATINSNDLTRIGNYTDYAVRWYNTGLE 216  
QY 292 -----VYFFRLVDPRRNVACIEHALEKLYIYKECCYD 324  
Db 217 RVWGPDSRDWVRYNQFRRELTILTVLIDVALFPNVDSPRYPIRTVSQLTREIY-----TN 270  
QY 325 PVRLWTEQYDGLK-----RQPKSP-----SITL-----DGLVY--VRRVLVTPCK 366  
Db 271 PV---LENFDGSRGSAQGIERSIRSPHMDILNSITITYDAHRGYVYWSGHQIMASP-- 325  
QY 367 VYFCGPEVNVNRLNRYSEDIDNLFVSVFDEWEKLYST--DLLPK-----ASTGSGVR 420  
Db 326 VGFSGPETF-----PLYGTMGNAAPQQRIVAQLGQGV- 358  
QY 421 TNYIERILSTL-RKGFVIGDKKFEFLAFSSQLR--DNSVMFASRPGTLA----- 468  
Db 359 ---YRTLSSLTLYRRPFNIG-----INNQLSVLDGTEFAYGTSNLPASVYRKSGTV 407  
QY 469 -----NDIRAWMGDFSOIKNVAKYAAALGQSGSSRETLSVLR----- 506  
Db 408 DSLDEIPQNNNVPPRQGFGRKLSVMSFRS-----GFSNSSVSIIRAPMFSWTHRSAT 461  
QY 507 -----HEIEVIPDKVH---GFSYVFSQDI--GKI-----SGDFAHRVASKGLOYT 548  
Db 462 PNTIDPERITQIPLVKAHTLQSGTIVYRGPGFTGDLRRTSGGPPAYTIVNINGQLQP 521  
QY 549 PSAFOIRGG-----YKGVGV-----DPDSSMKLSLRKSMKY----- 582  
Db 522 RYRARIYASTNLRIYTVVAGERIFAGQFNKMTDGTGDPPLTFQSFSYATINTAFTFPMQ 581  
QY 583 ESDNKLQVLGWSKYQPCYLNAR-----QLITLLSLTIGVKDE 618  
Db 582 SSFTVGAOT--PSSGNEYVIDREFLIPYATFEAYDLERAQKAVNALFTSINGIKTD 639  
QY 619 VLEQKQKRAVDOLDI-----LHDSLKAQBALEMLSPGENTNLIKAMLCNGYKPDPEPFLSM 675  
Db 640 VTDYH-----IDQVSNLVCLDSFCLDEKRELSEKVKHAKRLSDERNLLQDPNFKGINRQ 695  
QY 676 MLOTFRASKLLDLRFRIFIPNGTMMGCLEDES--RTLEYGOV-----FVQGTGAGH 727  
Db 696 LDRGWGRTDITIQRGDDVFKNYVTLPGTDECYPTLYQKIDESKLKATRIQLRGYI 755

QY 728 EESDHLHPFNNSRSTNSNFILKGNVVVAKNPKLPGCDIRVLKAVNVRALHHMVDVVFPQ 787  
Db 756 EDSQDL-----EYILIRYN---AKH-----ETVNPVGTSLWPLSAQSP 791  
QY 788 KGRPHPHNECS-----GSDLD 803  
Db 792 IGKCGEPNRCAPHLEWNPDL 812

## RESULT 13

US-09-873-873-28  
; Sequence 28, Application US/09873873  
; Patent No. US20020064865A1  
; GENERAL INFORMATION:  
; APPLICANT: Malvar, Thomas  
; APPLICANT: Gilmer, Amy Jelen  
; TITLE OF INVENTION: Polynucleotide Compositions Encoding Broad-Spectrum S-Endotoxin  
; FILE REFERENCE: MECO:210--2  
; CURRENT APPLICATION NUMBER: US/09/873,873  
; CURRENT FILING DATE: 2001-08-20  
; PRIOR APPLICATION NUMBER: US 09/253,341  
; PRIOR FILING DATE: 1999-02-19  
; PRIOR APPLICATION NUMBER: US 08/922,505  
; PRIOR FILING DATE: 1997-09-03  
; PRIOR APPLICATION NUMBER: US 08/754,490  
; PRIOR FILING DATE: 1996-11-20  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 28  
; LENGTH: 1177  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Hybrid Delta-Endotoxin  
US-09-873-873-28

Query Match 1.8%; Score 103; DB 10; Length 1177;

Best Local Similarity 18.2%; Pred. No. 2;  
Matches 168; Conservative 112; Mismatches 301; Indels 340; Gaps 45;

QY 102 FCGQISDDKFAVLGSTEVS-----IQFGIGLKKFFFLSSGSADYKQLQLSYENIMQVVLH 156  
Db 13 YNC-LSNPEVEVLGGRIETGYTPIDISLSLTQFLSEFVPGAGFVLGL-VDIIWGI--- 67  
QY 157 RPYG--QNAQFLL-----IQFGAPRIYKRL--NSCYSEFKETPDQWVRTTDFP 203  
Db 68 --FGSQWDAFLVQIEQLINQRIEFAFNQAIISRLGSLNYQIYAES-----FREWEAD 120  
QY 204 PSWIGLSSSLCLOFRGVRPLNPFEEFHYAERENNTITQGTFFFSQKSAVPNVQPP 263  
Db 121 PTNPALREEMRIQF-----NDMNSALTTPALFA-----VQNVQVP 156  
QY 264 -----EGISIPYKILFKISSLVQHGCCIPGALN----- 291  
Db 157 LLSVYVQAANLHLSVLDRVSVFGQWGFDAATINSNDLTRIGNYTDYAVRWYNTGLE 216  
QY 292 -----VYFFRLVDPRRNVACIEHALEKLYIYKECCYD 324  
Db 217 RVWGPDSRDWVRYNQFRRELTILTVLIDVALFPNVDSPRYPIRTVSQLTREIY-----TN 270  
QY 325 PVRLWTEQYDGLK-----RQPKSP-----SITL-----DGLVY--VRRVLVTPCK 366  
Db 271 PV---LENFDGSRGSAQGIERSIRSPHMDILNSITITYDAHRGYVYWSGHQIMASP-- 325  
QY 367 VYFCGPEVNVNRLNRYSEDIDNLFVSVFDEWEKLYST--DLLPK-----ASTGSGVR 420  
Db 326 VGFSGPETF-----PLYGTMGNAAPQQRIVAQLGQGV- 358  
QY 421 TNYIERILSTL-RKGFVIGDKKFEFLAFSSQLR--DNSVMFASRPGTLA----- 468  
Db 359 ---YRTLSSLTLYRRPFNIG-----INNQLSVLDGTEFAYGTSNLPASVYRKSGTV 407



```
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801.368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 370
; LENGTH: 1331
; TYPE: PAT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-370
```

```
Query Match 1.7%; Score 102; DB 10; Length 1331;
Best Local Similarity 18.8%; Pred. No. 3;
Matches 228; Conservative 170; Mismatches 413; Indels 404; Gaps 69;

Qy 44 SRAFAKOVADNIS---ADKIITIANRLXFGS---SVLKAW-EMKTDIV-----QLRAYVD 93
Db 251 NRYHSVSVWPNLKNMKYDK2IT---NTAFTARIDVMSNMLNFKTNLTIRRQELDDWIN 307
Qy 94 QMDGTLNFGQISDD-----KFAVLGS---TEVSIQFGIGLKKFFFLSSGSADYKL 143
Db 308 RPSISSNDCQEDFGVQWNCNCKMILAEOLMEKNIE-SIFOKKIFYPUSPMW--FKL 364
Qy 144 QLSY----ENIWQVVLHRYPGNQAOFLIQLFGAPRIY-----KRLNSCYFFFKETP 192
Db 365 KLHFIVYRETLTKMKIKYPERLSRLSLLAPVYLKIVELTSLYARKLKNPTMMIDQMI 424
Qy 193 DDQWRTTDFPSPMIGLSSSLCLFRGV-RLP---NFESEFFHYARENNITLOTGTFP 248
Db 425 DDF-----NAFIRLSVQKYTLTKYCSNLPFDVDFDPTF-----ENTVIEAIRYLE 470
Qy 249 FV-----SOKSALVPNVQPEGISIPYKILFKISSLVQHG---IPGPALNVYFFRLV 298
Db 471 FLNLKLIDSSKQ---NFKAPD-LLLYWDHLKNTGHYINGAETVIPNEFLKL-TLRV 524
Qy 299 DRRNRVACIHALEKLYIKECCYDV-----RWLTEQYD--GYLKGQRPKPSPI 348
Db 525 -----HKLQ-FYLLKQONFPPTFANASEAKWLSSIFENLGAMK-RKLNRFSNI 571
Qy 349 -----TLDDG---LVVRRVLVTPCKVYFCGPEV-----NYSNR 379
Db 572 LVKAFQNSAVYQINHNALVKLKDHAHYLVYSGNTFESSGVYNFAAPELLGCDNDTILR 631
Qy 380 VLNRVSEDIDNFRVSVFDEWEKLYSTDLLPKASTGSGVETNIYERILSLRKGFVIG- 438
Db 632 ILRN-----KSIGDLVPKLDIGNNL--NVYDITTKETDLNILVSK 670
Qy 439 --DKK---FEFLAFSSQL-----RQNSVW----- 458
Db 671 GEDSGIPYRVANSSDLDORHAHQSKKNFSTDPDQHLDEKNNEVFELEVALSSLGA 730
Qy 459 MFASRPGLTANDIRAWMGDFSQIKNVAKYAAR---LGQSGSSRETLSVLRIHEIETIPDV 515
Db 731 LVVLYPG---EPVVWDGVPVKLPGNLNFASNEMDLGKI--GNPNTLILLN----- 775
Qy 516 KVHGTSYVFSGIGKIS---GDFHRVASKCGLQYTPSAFQ-----IRY---GGYKG 561
Db 776 --QGSNYALTQIDKFNQTVGDSVSFIEKRCSLNSIESLQKINKAYKLYTYVLNNYK 833
Qy 562 VVGVDPDDSMK-----LSLRKSMKSYESDN-----IKLDVLGS 595
Db 834 ILG-----SFMKQCPGNELLNLSIFMGDFGRSFLKYNAFSSKRKYVIFLWYKLG-MNWL 888
Qy 596 K-----YQPCYLNRLITLLST-----LGVKDEVLEQKQKAEVQQLDAILH 636
Db 889 KFLVECDPTDQRTFRWCVLANDFAMQMTSGYNILALNVK-QFQELKERVSCVMSLLISH 947
```

```
Qy 637 DSLKAQAEALMSPGENTNLIKAMLCGYKPDAPFELSMMLQTFRASKLIDLTLRTRIFI 696
Db 948 FDMVGARATE---AEN-GMQQARLNDTEENIDEETLEINS--RLRLEAINTLEKTKM 1000
Qy 697 PNGRTMMGLDSERTLEYGOVFQFTGAGHGEFSDDLHPNNRSRSTNSNFI---LKGNVV 753
Db 1001 RNPRQMGKVLDAT---DQGNKYLLSLASSLSNVS-----MRWQKRSFISGGTGGQVY 1049
Qy 754 VAKNPCILHPGDIRVLKAVNVRALHVMVDCVVPQKGR-----PFPN--ECSGSDLDG 804
Db 1050 SAIN--LENGEILLAVKEIKHDTTMMKK--IFPLKEEMTVLEMLNHPNIVQYGVVHR 1105
Qy 805 DIYFVCDQDMIPPRQVQPMYPPAPSIQ--LDHVDVIEVEEYFTNYVNDLSGLIANA 862
Db 1106 DKYNIF-----MEYCEGGSASLLDHGRIDEMVTQVYTFELLEGILAYLHQS 1152
Qy 863 HVVFADREPDMAMSDPCKKLAEFLSTAVDPKTVGPAETPSQLRPKEYPDPMKPKDTSY 922
Db 1153 GVVRDRIKPE-----NILLDF--NGI-----IKYVDF-----G 1178
Qy 923 ISERVIGKLPKVKDKAPQASSIATFTROVARRSYDADMEVDGFEDYIDEAFDYKTEYDN 982
Db 1179 TARTVVGSRTRTVRNAVQDFGVETKSLN-----EMMGTPMYM----- 1216
Qy 983 KLGNLMDYYGIKTEAEITLGGIMKASKTFDRRKDAEALISVAVRALRKEARAFKRRNDID 1042
Db 1217 -----APETISGSVKGKLGAD---DVALGCVVLEMATGRRPWSNLDNE-- 1258
Qy 1043 DMLPKASAW---YHV 1054
Db 1259 -----WAIMVHV 1265
```

Search completed: November 6, 2002, 03:47:45  
Job time : 22 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:43:53 ; Search time 25 seconds  
(without alignments)  
4283.750 Million cell updates/sec

Title: US-09-782-874-2

Perfect score: 5864

Sequence: 1 MGKTIQVFGPPYLLSAEYWK.....RPVLNLSLRAQLSHRLVLK 1114

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9613422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1.\*

2: p1r2.\*

3: p1r3.\*

4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	5864	100.0	1114	2 T30819	RNA-directed RNA p
2	5070	86.5	1116	2 T30828	probable RNA-direc
3	1995.5	34.0	1133	2 T01920	probable RNA-direc
4	948	16.2	1215	2 T11660	hypothetical prote
5	932	15.9	1780	2 T20695	hypothetical prote
6	896.5	15.3	3228	2 T21381	hypothetical prote
7	362	6.2	966	2 F84582	hypothetical prote
8	350.5	6.0	929	2 H84582	hypothetical prote
9	335.5	5.7	905	2 G84582	hypothetical prote
10	222.5	3.8	604	2 F87936	protein M01G12.12
11	222.5	3.8	604	2 T23669	hypothetical prote
12	138	2.4	1021	2 T40328	conserved hypothet
13	133.5	2.3	2364	2 T40884	cytotoxin L - Clos
14	133	2.3	2810	2 T22928	hypothetical prote
15	131	2.2	463	2 S37872	low-temperature vi
16	130.5	2.2	1091	2 F83928	hypothetical prote
17	129.5	2.2	849	2 H86430	T518.2 protein - A
18	129	2.2	1733	1 RNBX2L	DNA-directed RNA p
19	128.5	2.2	839	2 T12827	hypothetical prote
20	126.5	2.2	1610	2 T11681	hypothetical prote
21	124	2.1	551	2 AC3572	lysine-tRNA ligase
22	123.5	2.1	1038	2 AF3286	ATP-dependent DNA
23	121.5	2.1	1745	2 T48366	anthranilate phosph
24	121.5	2.1	1972	2 S68176	TOG protein - huma
25	118.5	2.0	404	2 T21251	hypothetical prote
26	118.5	2.0	503	1 DECHE	glutamate dehydrog
27	117	2.0	688	2 S05380	hypothetical prote
28	116.5	2.0	1200	2 T15921	hypothetical prote
29	116.5	2.0	2137	2 T05244	hypothetical prote

30 116 2.0 968 2 C64691 type III restricti  
31 115.5 2.0 4436 2 E71086 hypothetical prote  
32 115 2.0 1163 2 E71481 probable swi/snf h  
33 115 2.0 4589 2 T14914 dynein beta heavy  
34 114.5 2.0 776 2 A11420 gamma-glutamylcyst  
35 114.5 2.0 2214 1 QZBYU2 pyrimidine synthes  
36 114.5 2.0 3191 2 T22945 hypothetical prote  
37 114 1.9 972 2 C71826 hypothetical prote  
38 114 1.9 1160 2 T39898 carbamoyl-phosphat  
39 114 1.9 5107 2 T29144 partial CDS - Caen  
40 113.5 1.9 633 2 T31160 maturase-related p  
41 113 1.9 744 2 A81719 exodeoxyribonuclea  
42 113 1.9 1748 2 S63127 probable membrane  
43 112.5 1.9 1188 2 E89896 chromosome segrega  
44 112.5 1.9 1816 1 S68960 laminin alpha4 ch  
45 112 1.9 989 2 T03838 telomerase catalyt

#### ALIGNMENTS

##### RESULT 1

T30819

RNA-directed RNA polymerase (EC 2.7.7.48) - tomato

C:Species: Lycopersicon esculentum (tomato)

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 21-Jul-2000

C:Accession: T30819

R:Schiebel, W.; Pellissier, T.; Riedel, L.; Thalmair, S.; Schiebel, R.; Kempe, D.; Loi

Plant Cell 10, 2087-2102, 1998

A:Title: Isolation of an RNA-directed RNA polymerase-specific cDNA clone from tomato.

A:Reference number: Z20885; M01D:99055198; PMID:9836747

A:Accession: T30819

A:Status: preliminary;

A:Molecule type: mRNA

A:Residues: 1-1114 <SCH>

A:Cross-references: EMBL:Y10403; NID:g4038591; PIDN:CAA71421.1; PID:g4038592

C:Genetics:

C:Note: RORP

C:Superfamily: Arabidopsis probable RNA-directed RNA polymerase

C:Keywords: nucleotidyltransferase

Query Match 100.0%; Score 5864; DB 2; Length 1114;  
Best Local Similarity 100.0%; Pred No. 0;  
Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKTIQVFGPPYLLSAEYVKSFLEKTYGTVCALVKQSGKGSRAFAKVQFADNISADK 60

Db 1 MGKTIQVFGPPYLLSAEYVKSFLEKTYGTVCALVKQSGKGSRAFAKVQFADNISADK 60

QY 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVQMDGITLNFCCOISDDKFAVLGSTEVS 120

Db 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVQMDGITLNFCCOISDDKFAVLGSTEVS 120

QY 121 IQFGIGLKFFPFFLSSGSADYKIQLSYENIQQVWVLRHPYQNAQFLLIQIFGAPRIYKRL 180

Db 121 IQFGIGLKFFPFFLSSGSADYKIQLSYENIQQVWVLRHPYQNAQFLLIQIFGAPRIYKRL 180

QY 181 ENSCYFFRETDDQWVTRTDPFPPSWIGLSSSLCLOFRRGVRLPNFEESFFHYAERENNI 240

Db 181 ENSCYFFRETDDQWVTRTDPFPPSWIGLSSSLCLOFRRGVRLPNFEESFFHYAERENNI 240

QY 241 TLGTGTFVFSQKSAVPNVQPEGISIPYKILFKISSLVQHGICIPGALNVYFFRLVDP 300

Db 241 TLGTGTFVFSQKSAVPNVQPEGISIPYKILFKISSLVQHGICIPGALNVYFFRLVDP 300

QY 301 RRRNVAICIEHALKLYIIECCYDVPVRWLTQYDGLKGRQPPKSPSIILDDGLVYVRV 360

Db 301 RRRNVAICIEHALKLYIIECCYDVPVRWLTQYDGLKGRQPPKSPSIILDDGLVYVRV 360

QY 361 LVTPCKVYFCGPEVNVNRLNRSYEDINFLRVSVDEWEKLYSTDLLPKASTGSGYR 420

Db 361 LVTPCKVYFCGPEVNVNRLNRSYEDINFLRVSVDEWEKLYSTDLLPKASTGSGYR 420

```
QY 421 TNYIERILSTRKGFVIGDKKFEELAPSSQLRDNSSVMFASRPGLTANDIRAWMGDFSQ 480
D 421 TNYIERILSTRKGFVIGDKKFEELAPSSQLRDNSSVMFASRPGLTANDIRAWMGDFSQ 480
QY 481 IKNVAKAARLGSGSGSRRTLSVLRHEIEIVPDVKVHGTSYVFSGIGKISGDFAHRA 540
D 481 IKNVAKAARLGSGSGSRRTLSVLRHEIEIVPDVKVHGTSYVFSGIGKISGDFAHRA 540
QY 541 SKCGLQYTPSAFAQIRYGGYGVGVDPDSSMKLSLRKSMKSYESDNIKLDVLGWSKYQPC 600
D 541 SKCGLQYTPSAFAQIRYGGYGVGVDPDSSMKLSLRKSMKSYESDNIKLDVLGWSKYQPC 600
QY 601 YLNPQLITLSTLCVKDEVLEQKKEAVDQDLDAIHDSLKAQALELMSPGENTNLIKAM 660
D 601 YLNPQLITLSTLCVKDEVLEQKKEAVDQDLDAIHDSLKAQALELMSPGENTNLIKAM 660
QY 661 LNCGYKPDAPFPLSMMLQTFRASKLLDLRTSRIFIPNGRTMGCDESRTLEYGVQFVQ 720
D 661 LNCGYKPDAPFPLSMMLQTFRASKLLDLRTSRIFIPNGRTMGCDESRTLEYGVQFVQ 720
QY 721 FTGAGHGEFSDLLHPFNNSRSTNSNFIKGNVYVAKNPCLHPGDIRVLKAVNRALEHMY 780
D 721 FTGAGHGEFSDLLHPFNNSRSTNSNFIKGNVYVAKNPCLHPGDIRVLKAVNRALEHMY 780
QY 781 DCVVFQKGRPHNECSGSDLDGDIYFVCWQDQDMPROVQPMEXPPAPSIQLDHDTI 840
D 781 DCVVFQKGRPHNECSGSDLDGDIYFVCWQDQDMPROVQPMEXPPAPSIQLDHDTI 840
QY 841 BEVEEYTNITVNDSLGIIANAHVVFADREPDMAMSDPCKKLAELFSIAVDFPKTGYPAE 900
D 841 BEVEEYTNITVNDSLGIIANAHVVFADREPDMAMSDPCKKLAELFSIAVDFPKTGYPAE 900
QY 901 IPSQLRKEVPDFMDKPKTSYISERVIGKLFKRVKDKAPQASSIAETRDVARRSDAD 960
D 901 IPSQLRKEVPDFMDKPKTSYISERVIGKLFKRVKDKAPQASSIAETRDVARRSDAD 960
QY 961 MEVDGFEDYIDEAFDYKTEYDNKLGNDYIGIKTEAELISGGIMKASKTFDRKKDAEI 1020
D 961 MEVDGFEDYIDEAFDYKTEYDNKLGNDYIGIKTEAELISGGIMKASKTFDRKKDAEI 1020
QY 1021 SVAVRALKEARAFKRRNDIDMLPKASAWYHYTYHTYGCYNQGLKRAHFTSPWCV 1080
D 1021 SVAVRALKEARAFKRRNDIDMLPKASAWYHYTYHTYGCYNQGLKRAHFTSPWCV 1080
QY 1081 YDQLTIQIKKARKARNPVLNLSLRAQLSHRLVK 1114
D 1081 YDQLTIQIKKARKARNPVLNLSLRAQLSHRLVK 1114

RESULT 2
T30828
RNA-directed RNA polymerase - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 20-Jun-2000
C:Accession: T30828
R:Schiebel, W.; Pelissier, T.; Riedel, L.; Thalmair, S.; Schiebel, R.; Lotsep
submitted to the EMBL Data Library, October 1998
A:Description: Isolation of an RNA-directed RNA polymerase-specific cDNA clone from Toma
A:Reference number: Z20890
A:Accession: T30828
A:Molecule type: mRNA
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Residues: 1-1116 <SCH>
A:Cross-references: EMBL:AJ011576; PIDN:CAA09697.1
C:Genetics:
A:Note: RdRP
C:Superfamily: Arabidopsis probable RNA-directed RNA polymerase

Query Match 86.5%; Score 5070; DB 2; Length 1116;
Best Local Similarity 86.08; Pred. No. 0;
Matches 959; Conservative 67; Mismatches 87; Indels 2; Gaps 1;

QY 1 MGKTIQVFGFPYLLSAEYVKSFLKNGTNGTYALEVQSKGRRAFRAKQVFNANKSAEF 60
```

```
D 1 MGKTIQVFGFPYLLSAEYVKSFLKNGTNGTYALEVQSKGRRAFRAKQVFNANKSAEF 60
QY 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDMDGTTLNFGCOISDDKPAVLGSTEVS 120
D 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDMDGTTLNFGCOISDDKPAVLGSTEVS 120
QY 61 IIDLASKGLYGPSPYLKAWEMKTDIVQPTTYVHMDGVTNLNFGCOISDKKPAVLGSKDVS 120
D 61 IIDLASKGLYGPSPYLKAWEMKTDIVQPTTYVHMDGVTNLNFGCOISDKKPAVLGSKDVS 120
QY 121 IQFGIGLKKFFFLSGGSADYKQLQSYENIWQVILHRPYGQNAQFLLQLFGAPRIYKRL 180
D 121 IQFGIGLKKFFFLSGGSADYKQLQSYENIWQVILHRPYGQNAQFLLQLFGAPRIYKRL 180
QY 121 IKFGIGLKKIYFFLSHASADYKQLQSYENIWQVILHRPYGQNAQFLLQLFGAPRIYKRL 180
D 121 IKFGIGLKKIYFFLSHASADYKQLQSYENIWQVILHRPYGQNAQFLLQLFGAPRIYKRL 180
QY 181 ENSCYSFKEKTPDDQWVRTDPPSWIGLSSSLCQLQFRRGVLPNFESEFFSYAERENNI 240
D 181 ENSCYSFKEKTPDDQWVRTDPPSWIGLSSSLCQLQFRRGVLPNFESEFFSYAERENNI 240
QY 181 EDCSCYFKEKTPDDQWVRTDPPSWIGLSSSLCQLFRRGVLPNFESEFFSYKESMNOF 240
D 181 EDCSCYFKEKTPDDQWVRTDPPSWIGLSSSLCQLFRRGVLPNFESEFFSYKESMNOF 240
QY 241 TLQGTFTFVSQKSLVNPVQPEGISIPYKILFKISSLVHQHCIPGALNVPYFRFLVDP 300
D 241 TLQGTFTFVSQKSLVNPVQPEGISIPYKILFKISSLVHQHCIPGALNVPYFRFLVDP 300
QY 241 ILQGTFTFVSQKSLVNPVQPEGISIPYKILFKISSLVHQHCIPGALNVPYFRFLVDP 300
D 241 ILQGTFTFVSQKSLVNPVQPEGISIPYKILFKISSLVHQHCIPGALNVPYFRFLVDP 300
QY 301 RRRNVACIEHALEKLYIKECCYDPVRMLTEQCYLKGROPKSPSITLDDGLVYVRRV 360
D 301 RRRNVACIEHALEKLYIKECCYDPVRMLTEQCYLKGROPKSPSITLDDGLVYVRRV 360
QY 301 RRRNVACIEHALEKLYIKECCYDPVRMLTEQCYLKGROPKSPSITLDDGLVYVRRV 360
D 301 RRRNVACIEHALEKLYIKECCYDPVRMLTEQCYLKGROPKSPSITLDDGLVYVRRV 360
QY 361 LVTQCKVYFCGPEVNVSNRVLNYSIEDINFLRVSFVDEWEKIHSTDLPRASTNGSTR 420
D 361 LVTQCKVYFCGPEVNVSNRVLNYSIEDINFLRVSFVDEWEKIHSTDLPRASTNGSTR 420
QY 421 TNYIERILSTRKGFVIGDKKFEELAPSSQLRDNSSVMFASRPGLTANDIRAWMGDFSQ 480
D 421 TNYIERILSTRKGFVIGDKKFEELAPSSQLRDNSSVMFASRPGLTANDIRAWMGDFSQ 480
QY 481 IKNVAKAARLGSGSGSRRTLSVLRHEIEIVPDV--KVHGTSYVFSGIGKISGDFAH 538
D 481 IKNVAKAARLGSGSGSRRTLSVLRHEIEIVPDV--KVHGTSYVFSGIGKISGDFAH 538
QY 481 IRNVAKAARLGSGSGSRRTLSVGRHEVEIPDVACSLHGTNYIFSDGIGKISADFAHR 540
D 481 IRNVAKAARLGSGSGSRRTLSVGRHEVEIPDVACSLHGTNYIFSDGIGKISADFAHR 540
QY 539 VASKCGLQYTPSAFAQIRYGGYGVGVDPDSSMKLSLRKSMKSYESDNIKLDVLGWSKYQ 598
D 539 VASKCGLQYTPSAFAQIRYGGYGVGVDPDSSMKLSLRKSMKSYESDNIKLDVLGWSKYQ 598
QY 541 VTIKCGLQYTPSFQIRYGGYGVGVAVDPYSSMKLSLRKSMKSYESDNIKLDVLGWSKYQ 600
D 541 VTIKCGLQYTPSFQIRYGGYGVGVAVDPYSSMKLSLRKSMKSYESDNIKLDVLGWSKYQ 600
QY 599 PCVNLNQLITLSTLCVKDEVLEQKKEAVDQDLDAIHDSLKAQALELMSPGENTNLIK 658
D 599 PCVNLNQLITLSTLCVKDEVLEQKKEAVDQDLDAIHDSLKAQALELMSPGENTNLIK 658
QY 601 PCVNLNQLITLSTLCVKDEVLEQKKEAVDQDLDAIHDSLKAQALELMSPGENTNLIK 660
D 601 PCVNLNQLITLSTLCVKDEVLEQKKEAVDQDLDAIHDSLKAQALELMSPGENTNLIK 660
QY 659 AMLNCGYKPDAPFPLSMMLQTFRASKLLDLRTSRIFIPNGRTMGCDESRTLEYGVQF 718
D 659 AMLNCGYKPDAPFPLSMMLQTFRASKLLDLRTSRIFIPNGRTMGCDESRTLEYGVQF 718
QY 661 EMLNCGYKPDAPFPLSMMLQTFRASKLLDLRTSRIFIPNGRTMGCDESRTLEYGVQF 720
D 661 EMLNCGYKPDAPFPLSMMLQTFRASKLLDLRTSRIFIPNGRTMGCDESRTLEYGVQF 720
QY 719 VQFTGAGHGEFSDLLHPFNNSRSTNSNFIKGNVYVAKNPCLHPGDIRVLKAVNRALEH 778
D 719 VQFTGAGHGEFSDLLHPFNNSRSTNSNFIKGNVYVAKNPCLHPGDIRVLKAVNRALEH 778
QY 721 VQFSGAGRQPFEEHSPFNDSGANSNCFILKGNVYVAKNPCLHPGDIRVLKAVNRALEH 780
D 721 VQFSGAGRQPFEEHSPFNDSGANSNCFILKGNVYVAKNPCLHPGDIRVLKAVNRALEH 780
QY 779 MDCVVFVPOKGRPHNECSGSDLDGDIYFVCWQDQDMPROVQPMEXPPAPSIQLDHDTV 838
D 779 MDCVVFVPOKGRPHNECSGSDLDGDIYFVCWQDQDMPROVQPMEXPPAPSIQLDHDTV 838
QY 781 MDCVVFVPOKGRPHNECSGSDLDGDIYFVCWQDQDMPROVQPMEXPPAPSIQLDHDTV 840
D 781 MDCVVFVPOKGRPHNECSGSDLDGDIYFVCWQDQDMPROVQPMEXPPAPSIQLDHDTV 840
QY 839 TIEVEEYTNITVNDSLGIIANAHVVFADREPDMAMSDPCKKLAELFSIAVDFPKTGY 898
D 839 TIEVEEYTNITVNDSLGIIANAHVVFADREPDMAMSDPCKKLAELFSIAVDFPKTGY 898
QY 841 TIEVEEYTNITVNDSLGIIANAHVVFADREPDMAMSDPCKKLAELFSIAVDFPKTGY 900
D 841 TIEVEEYTNITVNDSLGIIANAHVVFADREPDMAMSDPCKKLAELFSIAVDFPKTGY 900
QY 899 AEIPSQLRKEVPDFMDKPKTSYISERVIGKLFKRVKDKAPQASSIAETRDVARRSD 958
D 899 AEIPSQLRKEVPDFMDKPKTSYISERVIGKLFKRVKDKAPQASSIAETRDVARRSD 958
QY 901 AEIPSQLRKEVPDFMDKPKTSYISERVIGKLFKRVKDKAPQASSIAETRDVARRSD 960
D 901 AEIPSQLRKEVPDFMDKPKTSYISERVIGKLFKRVKDKAPQASSIAETRDVARRSD 960
QY 959 ADMEVDGFEDYIDEAFDYKTEYDNKLGNDYIGIKTEAELISGGIMKASKTFDRKKDAE 1018
D 959 ADMEVDGFEDYIDEAFDYKTEYDNKLGNDYIGIKTEAELISGGIMKASKTFDRKKDAE 1018
QY 961 SDMIVDGFEDYIDEAFDYKTEYDNKLGNDYIGIKTEAELISGGIMKASKTFDRKKDAE 1020
D 961 SDMIVDGFEDYIDEAFDYKTEYDNKLGNDYIGIKTEAELISGGIMKASKTFDRKKDAE 1020
QY 1019 AISVAVRALKEARAFKRRNDIDMLPKASAWYHYTYHTYGCYNQGLKRAHFTSPFW 1078
D 1019 AISVAVRALKEARAFKRRNDIDMLPKASAWYHYTYHTYGCYNQGLKRAHFTSPFW 1078
QY 1021 AIGVAVRALKEARAFKRRNDIDMLPKASAWYHYTYHTYGCYNQGLKRAHFTSPFW 1080
D 1021 AIGVAVRALKEARAFKRRNDIDMLPKASAWYHYTYHTYGCYNQGLKRAHFTSPFW 1080
QY 1079 CVYDQLTIQIKKARKARNPVLNLSLRAQLSHRLVK 1113
D 1079 CVYDQLTIQIKKARKARNPVLNLSLRAQLSHRLVK 1113
```

QY 451 QLRDNYMFASRPGLTANDIRAWMGDFSQIKNVAKYAARLGQFGSSRETLVLRHEIE 510

```

QY 355 VIVRVLVTPCKYFCGPEVNSNRVLKRNYSDDLNF LKVSFVDEWEKLSIDLLFRAAS 411
      ::::: ||| : : |||::: : ||| : || ::: |

```





```
Db 1091 NRPILKTGKVLITKNPCHVPGSVRVFVAVQPALAHLVDVVVFPQGRPRHPDENMAGSDL 1150
QY 803 DGDIVFVCDQDMIPROQVOPMEYPPAPSIQLDHVDVTEVEVEYTNVIVNDSGLIANA 862
Db 1151 DGEVSIINWQEMLDLYNEEAMVFPSSNAEDKEPTDDVVEFFRLYLQDSDIGRMSHA 1210
QY 863 HVFADREPDMAMSDPCKKLAEFLSIADVPFKTGPVPAEIPSQLRPKEY-PDFMDKPDKTS 921
Db 1211 HLAYADLHG--LFHENCHAIALKCAVAVDFPKSGVPAEPLSLSEFQECMTPTDYMMSGGKPM 1268
QY 922 YISERVIGKLFKVKDKAPQASSIATFTDVARRSYD-----ADMEV-DGPEYIDEAFD 975
Db 1269 YISTRUNGQLHKKAR-KVEEVLBEETRGVSFEYDKLICPEDVDVFFGNEIKLVQILT 1327
QY 976 YKTEYDNKLGNIIMDYIGKITEAEILSGGIMKASKTFDRKDK-----AEASIVAVRALR 1028
Db 1328 LRDEYVDRMQQLDEYDIEASVVSGHAAIKRLAGMERDDYSFHTDKVVELAYEKLY 1387
QY 1029 KEARAWF-----KRRDIDDMLPKASAWHYHTY-HTY---WCYNOG 1067
Db 1388 AVFRAKFFEFEGEENIENDKNTLKTAKMHEKIRQWVFVAYVQPRINKAGRCIGQ- 1446
QY 1068 LKRAHFISPPWCYDQILQIKK 1089
Db 1447 -----SLPWAWDALCDLRR 1461

RESULT 6
T21381
hypothetical protein F26A3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21381
R:McMurray, A.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19415
A:Accession: T21381
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3228 <WIL>
A:Cross-references: EMBL:Z78419; PIDN:CAB01701.1; GSPDB:GNO0019; CESP:F26A3.3
A:Experimental source: clone F26A3
C:Genetics:
A:Gene: CESP:F26A3.3
A:Map position: 1
A:Introns: 36/3; 91/3; 216/3; 263/3; 308/3; 702/3; 828/3; 866/1; 1013/3; 1171/3; 1231/3;
; 2443/1; 2595/3; 2663/1; 2753/3; 2813/3; 2859/1; 3052/3; 3136/1; 3176/1

Query Match 15.3%; Score 896.5; DB 2; Length 3228;
Best Local Similarity 24.8%; Pred. No. 6.7e-52;
Matches 314; Conservative 210; Mismatches 396; Indels 345; Gaps 50;

QY 69 LVFGS-----SYLKAWEM-----KTDT---VQLRAYVDQMDGTTLNFGC-QISDDKFA 112
Db 1732 IFGNLQGGTYNHNWVSFRSNISSRDMHLKHAEFFEDKDTMTVQFCEEKKQFE 1791
QY 113 VLGSTEVSIQFGLKKFFFLSSGADYKQLQSYENINWQVYLH-RPYGQNAQFLLIQLF 171
Db 1792 DSRKQVRV-----NYQLTIRDSIRRIIVDPVEGCT-CVHFEVN 1832
QY 172 GAPRIYK-LENSCYSFKEFTDDQVWRTDPPSW-----IGLSSSLCLO 216
Db 1833 CPLLIRKGYIDNKSFFKPF---YERQKRFCDWRNGNWNHGNPQDAIADSPFFTIE 1888
QY 217 FRGVGLPNFESFFHYAERENITITQGTFFVSQKSAVPNVQPPGESIPY----- 270
Db 1889 FKEIS-----TKEMYRVLRLSRTK-----VLIEFANLPSIDVPMGSHYPNRLNK 1937
QY 271 -----KILFKISSLVQHGCIQFPAI- 290
Db 1938 KSPDTSNAPIFREFLKEIFPPKYEIVDDKLIDVNEERKFSITYLIECLLSRGAIVKDQIL 1997
QY 291 -----NVYFER-----LVDPRRNRVACIEHALEKLYIYKECY 323
```

## RESULT 7

F84582

hypothetical protein Atg919910 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

```
Db 1998 LNEQKNWFLEIIWYRNNDNQLCEAALEDVLHLIDGRKR-IGSILKCLDKICOKRR-VM 2055
QY 324 DPVRLWTEQ--YDGYLKGROPPKSPSITLDDGLVYVRVLVTPCKVYFCGPEVNVSRVL 381
Db 2056 KLVNGLTEKESIEGYR-----VRKVFTPTFVYIAPETIMGRVL 2097
QY 382 RNYSEDIDNFRVSVFDEWEKLYSTDLLPKASTGSGVRNIYERILST-----LRKGFV 436
Db 2098 RFPKDGTRVLRVTFRDNKKM-----RSNVTGKLLDRTANKYLEHVR 2142
QY 437 IGDKKFEFLAFSSSOLRDNVW-----MEASRGLT-----ANDIRAWMGD 477
Db 2143 IANREYFLGCSNQRMDNGAIFWMRTDQKLDRFYKCNPTASNINPKKIDEVRFOLGR 2202
QY 478 FSOIKNAVAKYAAARLQSGFSGSRRETLVLRHE-----IEVIPDVKHGTSYVYFSDGIGKIS 532
Db 2203 FSEIENVKMLARLQGCFTQSRLTGLGRDDYCTVDLTGGRATNGSEYTFSDGVGMS 2262
QY 533 GDFAHRVASKGLQY---TPSAFOIRYGGYKGVGVDP-----DSSMKLS-- 574
Db 2263 YQFAQESQ---AMQFGKAVPSCFOIRFRNGKGVIAIEPFLDEIRKVALNVGVTSMKKAKC 2320
QY 575 -LRKSMKYESDNKLDVLGHSKYQPCY--LNRQLITLTLTGKDEVLEQKQKAVDQ 630
Db 2321 LFRPSQIKFOAKISGDOIEMVKFSSAVVALNKPFINL-----DQVSEMQSLDCHKR 2374
QY 631 LDAILHSLKAQBALELMSPGENT---NILKAM-----LNCGYKPDABEFLSMWL 677
Db 2375 ITSRIEELMDRQILSFAKQNMNEETFCRNKLEFFRRIDIDNLRITMGTUSSEFFRSLI 2434
QY 678 QT---FRASKLLDLRTSRIPN--GRTMMGCLDESRTELEYGVQVFTGAGHGEFSD 732
Db 2435 KASIKFSITKOL---CKEQIOLPSLGSMLGVDEYGRLOYGQIFVQYT---KNYKKK 2487
QY 733 LHPFNNSRSTNSFILKGNVVAKNPCLHPGDIRVLKAVNVRALHMHVDCVFFPKQKRP 792
Db 2488 LPPRDSNNKHVSGSEITVGTLLTKNPICVPGDVRIEAVDIPELHMKDCVVVFPQHGRP 2547
QY 793 HPNCSGSLDGDIDYFVCWDDMIPPROVQPMYPPAPSIQLDHVDVTEEV--BEYFTNY 850
Db 2548 HPDEMAGSDDLQGDYSYIWDQELLERNEEPPDF-AVEKIKVPYDREKLDVLMREFVYTY 2606
QY 851 TVNDSLGIITAHAVHVFADREPDMAMSDPCKKLAEFLSIADVPFKTGPVPAE----- 900
Db 2607 LKLDVSGVQISNHLNHSQYG--LNSRVCMDLAKKNCQAVDFTKSGOPDPLETKWRADP 2664
QY 901 -----IPSQLPKEYPDP-MDKPKTSYISERVIGKLF-----KVKDKAPQASSI 945
Db 2665 VTFEVIPEE-NPERIPDFHGNERSPMYSPRLCGKLFREFQAIDNVIKISEEDEQYNI 2723
QY 946 ATFTDRVARRSYDADMEYDGFEDYIDEAFDYKTEYDNKLGNDMDYYGKIKTEAELISGGM 1005
Db 2724 -----ELDETIFTVTGPERYMESAQKOLSSYNGQLRSIMENYIGIRSEIGMSGCIV 2773
QY 1006 K-ASKTFDRKD-----AEASIVAVRALRKEARWF----- 1035
Db 2774 EMRRISDKDDDDSYNTNMIMETKMTSLVCKFERIFFEFGFTVKCTLLPNAYDNGN 2833
QY 1036 ---KRRNDID-DMLPKASAWHYHTYHPTWGCYN--QGLKRAHFISPPWCYDQILQIKK 1089
Db 2834 CLNYRCEPDQEVKKA VAWYR-----ACYECAQSTREVRKLSFAWIAVDYIAKVK 2885
QY 1090 DKARN 1094
Db 2886 TNVLN 2890
```

C:Accession: F84582

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: F84582

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-966 <STO>

A:Cross-references: GB:AE002093; NID:g3687227; PIDN:AAC62125.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g19910

A:Map position: 2

Query Match 6.2%; Score 362; DB 2; Length 966;

Best Local Similarity 22.0%; Pred. No. 2.9e-16;

Matches 198; Conservative 136; Mismatches 306; Indels 262; Gaps 42;

QY 323 YDPVRW--LTEQYDYLKGRQPPSPSITLDDGLV-YVRRVLVTPCKVFCGPEVNSNR 379

DB 184 YEAAVNNLGRQYCPKERRRPLEG-----DSGMTHYYQCHVATDGSYKFKGHLENTGT 238

QY 380 VLNRYSIEDNPLRYSFVDEWEKLYSTDLLPKASTGSGVRTNIYERILSLRKGEVIGD 439

DB 239 HLHKVLGD-DNVLTVKFKVLGVETVCNDLY---STYKGIKLN-----GIMVGL 283

QY 440 KKEFLAF-----SSQLRDSNVMMFASRPGLTANDIRAW 474

DB 284 RRYEFKDGKKEKKDVKSTGKVCYFIRTDSTAFYDMQNPYIFA---GKSMHEARMH 340

QY 475 MGDFQIKNVAKYARLQSGSSRETLSVLRLHEIVPDVKVCTSVFSDGIGKLSGD 534

DB 341 FMHVNTLSLPMYMR-----YSIDDDKDVLDK---NGKPCIHSDGIGYISED 386

QY 535 FA-----HRVASK-----CGLYTPSAFQIR--YGGY--RG----- 561

DB 387 LARMCVPNIFKGMRSNNIQSKNLFEGQGPCG-QEPLLIIQPIFYNGYAVKGTFLT 445

QY 562 -----VVGVDPPDSNKLRLSKMSKYESDNKLDVLGWSK-YQPCYNLRQLITLLSLG 614

DB 446 KKLPPRTVQVRP-SMKIYVEDTLNLSFTN-SLEVVTTSNPPRKARLSRLVALLSYGG 503

QY 615 VKDEVLEQKQKAVQDLDAILHDSLKAQALELMSGENTNILKAMLCG---YKPD-- 669

DB 504 VPNDFFNLRLNTLEESKTIYSERAA-----FKAAIYGGDQYTADMIL 548

QY 670 -----EPF-----LSMQLQTPFRASKLLDLRTSRIFIPNGRTMGCLESRTLEYGVQV 719

DB 549 VGIPLDEPYLKRLLSYLLKTER-----NALKAGREPIDSEYIIMGTVDPTELKENEICV 603

QY 720 QFTGAGHGEFSDLLHPFNNSRTSNFNLKGNVVAKNPCPLHGPDIRVLKAVNVRALHMM 779

DB 604 -----ILHSGQ-----ISGDVLYVRNPLHFGDIHVLKATYVKALEDY 641

QY 780 VD-----CVVPFQKGRPHNECSGSDLDGDIYFVCDQDMIPPRVQPM----- 825

DB 642 VGNAPFAVFPQKGRSLGDEIAGGDFDGDYFISRNPKLL--EHFKPSEPVSSSKPSK 699

QY 826 -YPPAPSIOLDHDVTIEEVEYF--TNIVNDSLGIIANAHVVFADRPDMAMSDPK-- 880

DB 700 IYGRKPSLSSELEELFKMLKARCKRDVIGMAADCWLGIMD--PFLTGDSEAKE 757

QY 881 -----KLAELFSAIVDFPKTGPAPRIQSLRPKPEYDPMPKPKTSIYSERVIGKL 931

DB 758 KYERKNILKLDIYYDALDAPKGAKVLDLPDLLEIKNFPHPHYMERDKRDRFSTSLGLI 817

QY 932 FRKVKD---KAQOASSIAT---FTDVARRSYDADMEVDGDFEYDIDEAFDYKTEYD---- 981

DB 818 FDTVDSHNAEPPPSISKLTWFEDEPVPKSH-MDKFTSWENYRSEMSQAMMETDKYKR 876

QY 982 NKLGN-----LMDYYGINKTEAILSGGIMKASKTFDRRKDAEAI SVAVRALRKEARAWF 1035

DB 877 NQLTNEVIQRYKQDFY-----AAGFEDSNKS----- 903

QY 1036 KRNRDIDDLPLPKASAWHVHTYHTYMGYNOGLKRAHFISFPCWCVYDQLI---QIKKDKA 1092

DB 904 -----LEELYQALALYNNVYD---YATQEGYAKC---TFANVAGPVLCRFLAKTKD 951

QY 1093 RN 1094

DB 952 KS 953

RESULT 8

H84582

hypothetical protein At2g19930 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: H84582

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: H84582

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-929 <STO>

A:Cross-references: GB:AE002093; NID:g3687225; PIDN:AAC62123.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g19930

A:Map position: 2

Query Match 6.0%; Score 350.5; DB 2; Length 929;

Best Local Similarity 22.6%; Pred. No. 1.7e-15;

Matches 179; Conservative 128; Mismatches 279; Indels 207; Gaps 39;

QY 389 DNFLRVFVDEWEKLYSTDLLPKASTGSGVRTNIYERILSLRKGEVIGDKKFEFLAF- 447

DB 212 DNVLTVKADV--QKSSSTYSIDHYFTYKGIKLN-----GIMIGLRRYQFFVFK 258

QY 448 -----SSQLR-----DNSVMMFASRP---GLTANDIRAWGDFSQIKNVAK 486

DB 259 DGGKEKKKDLSTKKVKCYFIRTDSTAFYDMQNPYILTGSITYEARMHFHVHAPILAN 318

QY 487 YAARLQSGSSRRTLSVLRLHEI--EVIPDVKVHGT-----YVFDGIGKISG 533

DB 319 YMARFSLIL-SKTKLEVDMTGITFDQIDDIHCHDQDQKDVLDKNKKPCIHSDGTGYISE 377

QY 534 DFAHR-----VASKC-----GLQYTPSAFQIRYGGY--KVVGVDPPDSMK-LSLRKSMK 581

DB 378 DLARMCPNIFKGRKLSRESIQPLLIQIRMFYDGYAVKGTFLNKKLCPRTVQVRPSMIK 437

QY 582 YESDNIKLDVLGWSKYQPCYNLRQLITLLSTLGVKDEVLEQKQKAEAVDQLDALHDSLKA 641

DB 438 VKSD-----PSLSNFSTFNALFVITRSQIFGAA-----ALNYGEMDD 475

QY 642 QEALIELMSGENTNILKAMLCNGYKPDAPFLSMQLQTPFRASKLLDLRTSRIFIPNGRT 701

DB 476 QNAQAQMLIVGIPLD-----EPHLKNYLSILLKTERN---DLKA-GKLPVTESY 520

QY 702 MMGCLESRTLEYGVQVQFTGAGHGEFSDLLHPFNNSRTSNFNLKGNVVAKNPCPLH 761

DB 521 LMGVDTGALKEDVCVILES---GQIS-----GEVLYVRNPLH 558

QY 762 PGDIRVLKAVNVRALHMMVD---CVVPFQKGRPHNECSGSDLDGDIYFVCDQDMIP 817

DB 559 FGDTHILKATYVKALEEVYGVNKSFAVFPQKGRSLGDEIAGGDFDGDYFISRNPELL- 617

QY 818 PRVQPM-----YPP-----APSIOLDHDVTIEEVEYF--TNIVNDSLGIIANAH 863

DB 618 -ENFKPSEPVSUVTTPSKNSGRAPS--QLSPFEELEELFEMFLTAGFHASNVIGIADSW 675



Db 417 SIWKFHICQKRLVNLQNGNSEQ-----IEEGYQVRKVITFTRVYIYPPEMI 468  
QY 376 VSNRVLNRKSEDDNLFVRSVDDEWEKLYSTDLPLKASTGSGVTRNIYERILST-----430  
Db 469 MGNRVLNRFDGTHVLRVTFRDDNNRKM-----RANATGELLIDICVKKY 513  
QY 431 LRKGFVIGDKKFFLAFSSQLRDNDSVWMFAS-----RPGLTANDIRAW 474  
Db 514 LEHGIVANRDFGLGSSQMRDNGAYFMVKNTDNRHKNACKMNSKFKPNI--DSVRNQ 571  
QY 475 MGDFSQIKNVAKYAARLGOSFGSSR 499  
Db 572 LGNFLOIENIPKLMARLQGCTQSR 596  
RESULT 11  
T23669  
hypothetical protein M01G12.12 - Caenorhabditis elegans (fragment)  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T23669  
R:Lennard, N.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19779  
A:Accession: T23669  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-604 <WIL>  
A:Cross-references: EMBL:Z81571; PIDN: CAB04618.1; GSPDB: GN00019; CESP: M01G12.12  
A:Experimental source: clone M01G12  
C:Genetics:  
A:Gene: CESP: M01G12.12  
A:Map position: 1  
A:Introns: 35/3; 88/3; 183/3; 224/3; 273/3; 331/3; 420/3  
Query Match 3.8%; Score 222.5; DB 2; Length 604;  
Best Local Similarity 20.7%; Pred. No. 4.6e-07;  
Matches 117; Conservative 74; Mismatches 197; Indels 177; Gaps 20;  
QY 43 GSRAFAKVFADNITSADKIITLANNRLYFGS----SYLKAWEMKTDIVQLRA-----90  
Db 101 GSDFWAK-----TLDCYIEVPLAAMYFGNIQGYTIVISHWQISFGSKKISADANELLNK 153  
QY 91 -----YVDQMDGITLNFQCOISDDKFAVLGSTVSIQFGIGLAKFFFFLSSGSADYKLOL 145  
Db 154 IVAFEFEDRADMITVTQC-----LKKRINIQITII 184  
QY 146 SYENIWQVLRHPYQNAQFLLIQLFGAPRI-----YKLEN-SCYSFF 188  
Db 185 RKDTIRAIIVDPQVDMNKTIRHFEELNCPPLIROGSDVDDKPKSTQKPKYKTRNYSCIG-T 243  
QY 189 KETPDQWRRTDPPSPWIGLSSSLC-----LQPRGYRL--PNFEE- 228  
Db 244 KEYGSPHSAISDPFTTIELQKQSDGSDGSDNTLRYVLSRLSRGTGVOIEFANPKV 303  
QY 229 -----SRFHYAERENNTITLGTFTFVSQSALVPNPQPEGISIPYKILFKISSLV 280  
Db 304 DVPYGVYPLRYPTSEKSAFECFIYNCFPKMKIIDAOSINENDGRQFAIYILIECLL 363  
QY 281 QHGCIIP-----GPNLVN---YFFRLVDPRRNRWACIEHALEKLYIYE-----320  
Db 364 SRGAIVKDVQLTDBICWQGLGLITHYL-----ENDKLCEAALEDLIYLDGRKRIG 416  
QY 321 ---CCYDPV---RWLTEQYDGLKGRQPKSPSITLDGLVVRVRLVTPCKYVFCGPEVN 375  
Db 417 SIWKFHICQKRLVNLQNGNSEQ-----IEEGYQVRKVITFTRVYIYPPEMI 468  
QY 376 VSNRVLNRKSEDDNLFVRSVDDEWEKLYSTDLPLKASTGSGVTRNIYERILST-----430  
Db 469 MGNRVLNRFDGTHVLRVTFRDDNNRKM-----RANATGELLIDICVKKY 513  
QY 431 LRKGFVIGDKKFFLAFSSQLRDNDSVWMFAS-----RPGLTANDIRAW 474

Db 514 LEHGIVANRDFGLGSSQMRDNGAYFMVKNTDNRHKNACKMNSKFKPNI--DSVRNQ 571  
QY 475 MGDFSQIKNVAKYAARLGOSFGSSR 499  
Db 572 LGNFLOIENIPKLMARLQGCTQSR 596  
RESULT 12  
T40928  
conserved hypothetical protein SPCC132.01c [imported] - fission yeast (Schizosaccharo  
C:Alternate names: conserved hypothetical protein SPCC132.17c  
C:Species: Schizosaccharomyces pombe  
C:Date: 20-Oct-2000 #sequence\_revision 03-Nov-2000 #text\_change 03-Nov-2000  
C:Accession: T40928; T40948  
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Aert, R.; Voickaert, G.  
submitted to the EMBL Data Library, October 1999  
A:Reference number: Z21958  
A:Accession: T40928  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-549 <LYN>  
A:Cross-references: EMBL:AL121807; PIDN: CAB58128.1; GSPDB: GN00068; SPDB: SPCC132.01c  
A:Experimental source: strain 972h-; cosmid c132  
R:Lucas, M.; Gaillardin, C.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, January 1999  
A:Reference number: Z21959  
A:Accession: T40948  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 374-1021 <LUC>  
A:Cross-references: EMBL:AL035259; PIDN: CAA22870.1; GSPDB: GN00068; SPDB: SPCC132.17c  
A:Experimental source: strain 972h-; cosmid c1322  
C:Genetics:  
A:Gene: SPDB: SPCC132.01c; SPDB: SPCC132.17c  
A:Map position: 3  
Query Match 2.4%; Score 138; DB 2; Length 1021;  
Best Local Similarity 21.3%; Pred. No. 0.61;  
Matches 123; Conservative 93; Mismatches 215; Indels 146; Gaps 31;  
QY 567 PSSSKLSLRK-----SMKSYESDNILKDLVG-----WSKYQPCVL-----NRQ 605  
Db 69 PLSGFVTLRKHKISRRITGVSQLGTRVLTFTFGGANDQPDWYIYLVCEFFAAGNVL 128  
QY 606 LI-----TLSTLGV---KDEVLEQKQEAVDOLDAILHDSLKAQEALEMLSPGENTNIL 657  
Db 129 LLDGHYKILSLRVVTFDKQYVAVGVQKYNLDK--NLVNDN--KSQSTIPHMT--AERLNL 185  
QY 658 KAMLNGY-----KPDAPFLSMMLQTFRASKLLDLRTSRIFIPNGRTMMGCLDES 710  
Db 186 LDEISTAVASPTSIINEPLPDQQLSSSTKPIKVPKPVLSRKALTI-----R 230  
QY 711 TLEYGVQVFTGAGHGEFDDLHPFNNSRSTNSNFIKGNVVVAKNCPCLPGDIRVLKA 770  
Db 231 LGEYGNALIE-----HCLRSKLDPLFPACQLCADETKNDLLAA-----FQEAD--SILAA 280  
QY 771 VN-----VRALHHMVDVCFVFKGKRPHPNECSGLDGIYFVCDQDMIPRQVQP 823  
Db 281 VNKPPVKGYIFSLEQALTNAADPQ-----HPEECT-----TLY-----EDFHP---FQP 321  
QY 824 MEYPPAPSTQLDHDVTIEVEVEYFTNYIVNDSLIGITANAHVVFADREPDMASDPCKKLA 883  
Db 322 LQIQVANRKCMEFFTYNECVDEFFSS--IEAQKLRKRAHRLATAERLESQEDQARKLQ 380  
QY 884 ELFSIAVDFPKTGVPFAEIPSQLRPKEY---PDFMDPKDTSYISERV-----IGKLF 932  
Db 381 SL-----QDAQATCALRAQAIAEMNPVLEA--IISYINSLNGLNDWLDIEKLI 427  
QY 933 RKVKDKAPQASSIAFTTRVAR-----RSYD--ADMEVDFGFDYIDEAFDYKTEYDN 982  
Db 428 QSKRRSRPAAAIQIPLKLIKNAVTFLPNPESVDNSESSESTSDDDLDSDSD-----DN 482

QY 983 KL--GNLMDYVGIKTEAEILSGIMKASKTFDRR-----KDAEASVAVRALRKAR 1032  
DB 483 KYKEGVSSKF-IAVDELISLGAFAFNARKQYELREALIKETKTAASAKALKSTQRKIE 541  
QY 1033 ANFKRRN--DIDMLPKASAWYHVTHVPHYWCYNOG 1067  
DB 542 QDLKRSTTADTORILLGRKTFEKEFH---WFISSEG 575  
RESULT 13  
I40884  
cytotoxin L - Clostridium sordellii  
C:Species: Clostridium sordellii  
C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 15-Oct-1999  
C:Accession: I40884  
R:Green, G.A.; Schue, V.; Montell, H.  
Gene 161, 57-61, 1995  
A:Title: Cloning and characterization of the cytotoxin L-encoding gene of Clostridium sordellii  
A:Reference number: I40884; MUID:95369733; PMID:7642137  
A:Accession: I40884  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2364 <RES>  
A:Cross-references: EMBL:X82638; NID:g1000694; PIDN:CAA57959.1; PID:g1000695  
C:Superfamily: cpl repeat homology  
C:Keywords: cytotoxin

Query Match 2.3%; Score 133.5; DB 2; Length 2364;  
Best Local Similarity 18.9%; Pred. No. 4.7; Mismatches 365; Gaps 57;  
Matches 208; Conservative 142;

QY 47 FARVQPADN--ISADKIITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDQ-MDGTILNFG 103  
DB 189 YKKSQIEENPEFIIDNII-----KYLSEYK-----DLKALNKYIEESLNKITANN 237  
QY 104 CQI-----SDDKFAVLGSTEVIQIGLKKFFFLSSGSADYKQLQSYENIWQVVLHR 157  
DB 238 NDIRNLEKFADELDVRLYNOELVERNLAAASDILRMLKEDGYLVDVILPGI----- 293  
QY 158 PYGONAQFLLIQFGAPRIYKRENSCYSEFFKETPDQWRTTDFPPSWIGLSSSLCLOF 217  
DB 294 -----QPDLEKSINK-----PDS--ITNT-----SWEMIKLEAIMKY 323  
QY 218 RRGVRLPNFEESPHYAERENNITLGTFTFFVSQKSALVNPVQPEGISI---PKILF 274  
DB 324 KE--YIPGYTSKNFMDLDER---VQSFESALSRSKSEIPLPLDDIKVSPLEVKIAF 377  
QY 275 KISLVQHGCIQGPALNVYFRFLVDRRRNVACIEHALEKLYIKCECYDPYRWLWTEQYD 334  
DB 378 ANNSVINQALI-----SLKDSY-----CSDLV---INQIK 404  
QY 335 GYLKGROPPKSPSITLDDGLVYVRRVLVTPCKYVFGPEVNSVNRVLRNYSEDIDNLRV 394  
DB 405 NRYKILANDLNPSI--NEGTF-----NITMKI-FSDKLASINEDNMFMKIKITNLYKV 456  
QY 395 SFVDEWEKLYSTDLPLPKASTGGVTRNIYERILSLTRKGFVIGD-----KRF 442  
DB 457 GFAPD-----VRSTINL---SGPGVYTGAYQDIL-----MFKDNSTNIHLEPELRNF 501  
QY 443 EFLAFSSQRLD---NSVMWFASRPGLTANDIRAWMGDFSOIKNVAKYAARLQSGFSSR 499  
DB 502 EFPKTKISQTEQEITSLNSF-----NQARA-KSQFEYKK-GYFEGALGEDDNLDF 551  
QY 500 ETLVLRLHEI---EVIPDVKVHTSVYVSDGIGKISGDFAFHRVASKGL-----QYTPSAF 552  
DB 552 AQNTVLKDVYKSKILSSMKTRNEYIHY--IVQLQGDKISYEAS-CNLFSDKDPYSSILY 608  
QY 553 QIRYGGVKGVGVD-----POSSMKLSLRKSMKY-----ESDNIKLDVLGSKYQ 598  
DB 609 Q-----KNIEGSETAYYYVADAIE-----KEIDKTRIPYQSNRKNIKLTFIGHGK-- 655  
QY 599 PCYLNROLITLLSLGVLKDEVLEQKQKAEVDQDLDAIHLSDSKAQEALELMSPGENTNLIK 658

DB 656 -----SEFNTDTFANLDV---DSLSSE--IETI-----LNLAK 683  
QY 659 A-----MLNC---GYPDAPFPFLSMMLQTFRASKLLDLTRSRIFIPNGRTMGC 705  
DB 684 ADISPKYIEINLLGNCNMFYSISAE-----ETYPGKLLKIKDRVSELMPS-----I 730  
QY 706 LDESRTLEYGVFVQFTGAGHGEFSDDLHPENN-----SRSTNSNFIKGNVVVAK 756  
DB 731 SODSITVSANQYEVIRNEEGREILDHSGKWKINKEESIIKDISKEYISFPNPKENKIIVK 790  
QY 757 NPCLHP-----GDIRVLKAVNVRLHHMVDCVFPQKGRPHHPNECSGSLDGDGIYFCW 811  
DB 791 SKYLHELSTLLQELR-----NNANSSDIDLE----- 816  
QY 812 DODMIPPRQVQPMYPPAPSPAPSTQLDHDVTIEVEEYFTNYIVNDSLGIIANAHVFA---- 867  
DB 817 -----KKVMLTECEINVANSDIROIVEGRIEE--AKNLTSDSINYIKNEFKLIESISD 867  
QY 868 ---DREPDMAUSD-----PCKLAELFSAVDFPKTG-----VPAEIPSQLR---PK 908  
DB 868 SLYDLKHQGLDDSHFISFEDISKTENGFRIRFINKETGNSIFITEKEIFEYATHISK 927  
QY 909 EYDPFMDKPDKTSYISERVYKLFKVKDKAPQASSIATFTRDVARRSYDADMEVDGFED 968  
DB 928 EISNIKDT-----IFDNVNGKLVKKV-----NLDAAHEVNTLNS 961  
QY 969 --YIDEAFDYKTEYDNKGLNLMDDYGIKTEAEILSGGIMKASKTFDRRKDAEASVAVRA 1026  
DB 962 AFFIOSLIEYNTKES--LSNLSVAMKVQVYQALFSTGL---NTTDASKVVVELYSTAL-- 1015  
QY 1027 LKKEARAWFKRRNDIDDMPLPKAS 1049  
DB 1016 -----DETIDLPTLS 1026  
RESULT 14  
T22298  
hypothetical protein F46C3.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T22298  
R:Cottage, A.  
submitted to the EMBL Data Library, November 1995  
A:Reference number: Z19543  
A:Accession: T22298  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2810 <WIL>  
A:Cross-references: EMBL:Z66563; PIDN:CAA91469.1; GSPDB:GMO0028; CESP:F46C3.3  
A:Experimental source: Clone F46C3  
C:Genetics:  
A:Gene: CESP:F46C3.3  
A:Map position: X  
A:Introns: 67/2; 172/3; 204/2; 310/3; 366/2; 431/2; 684/3; 739/3; 786/3; 823/2; 880/1958/3; 1999/1; 2078/1; 2117/3; 2159/2; 2220/1; 2269/3; 2306/1; 2399/2; 2444/1; 2488/  
Query Match 2.3%; Score 133; DB 2; Length 2810;  
Best Local Similarity 17.7%; Pred. No. 6.7; Mismatches 344; Gaps 46;  
Matches 169; Conservative 141;

QY 252 OKSALVPNVQPP-----EGISIPYKILFKISSLVQHGCIPGPALN-----VYF 294  
DB 1998 RKEPLVP--QPPRPVVAEKPAVKFVAPWKLTIKIR-----KEMFYPGVNLNDIQDQV 2050  
QY 295 FRLVDPRRNVACIEHALEKLYIKCECYDPYRWLWTEQDGYLKGQPP---KSPSITL 350  
DB 2051 AQIVEDCKKSYK-----YRIREQDRKQVETVLRQYE-----IPPSDLNNSNIHP 2095  
QY 351 DGLVYVVRVLYTPCKVYF-----CGPEVNSN-----RVLRNYSIEDIDNLR 393  
DB 2096 DVKAVIELARLWP--LIFYNQYEVVEKRPDESVTIFAISEGILFIVHTPHDLENPLK 2153  
QY 394 VS-----FVDEWEKLYSTDLL-----PKASTGSGVTRNIYERILSLTRK---GFVI 437

Db 2154 IDDFPPTIADVLEANDILSVHVRHEDEENAYSAVRIKTNQAPQIKTKTIDRLCSGGVV 2213  
QY 438 GDKKEFLAFSSQLRDSNVMFASRPGLTANDIRAMWDFSQKNVAKYAAARLQGSFGS 497  
Db 2214 PKRKE-----VLARTGREPPVGNWL--YGKIEN--RFGFLLAQYVDS 2252  
QY 498 S-RETLVLRHEIEVIPDVKHGTSYVSFSDGIGKISGDFAHVASKCGLQVTPSAFQIRY 556  
Db 2253 TGDGNVPIRHTSDRDERVR-----FFDDEV----- 2280  
QY 557 GYGKGVGVDPSSMKLSLRKSMKY---ESDNIKDLVGLGSKYQPCVNLNQLI----- 607  
Db 2281 -----PFSSERYTMIDFATRYFKPKKKKQETAWMEDIS-----QIVREFSEKP 2324  
QY 608 ---TLLSLTGKDEVELEKQKEAVDQDAIL-----HDSLKAQEALEMLSPGENTNILKAM 660  
Db 2325 ISQSLADLG-----NEESYAVETFAIMKFMGDEPLKKSES-----TDVFKVL 2371  
QY 661 LNCGYKPA--EPFLSMMLQT-----FRASKLLDLRTRSRIFIPNGRMTMGCLD 707  
Db 2372 LICHQPTLRDEVCQLIKQITSNISQKPNLSALRAWLLTIT--AVFPSSLTK----- 2424  
QY 708 ESRILEYGVQVFTGAGHGFSDLLHFPNNSRSTNSNF-----I 747  
Db 2425 -----PYVLQYLGDNADENQRPFH--GTARICQTNMIQTFYGRKVLNNALEVQOI 2474  
QY 748 LKGNVY-----VAK-----NPLCHPGDI-----RVLKAVNVRALHH 778  
Db 2475 TVGSLVLYFCBSTDGCQLRQCAFYSKDHNVQSQTLRITVAEMIQLCNLLNVRSLHE 2534  
QY 779 MYDCVVFQKGRPHNCSGSDLDGDIYFVCWDQDMTPPRQ-----VQMEYPPAP 830  
Db 2535 QQEFSLCYTVGDKHLNCKNDNYLMDI--ITESHKLLPQFYLVKRTVWVHPLRYDNAA 2592  
QY 831 STQLDHDVTIEVEEYFTNYIVN--DSLGLIANAHVVFADREPDMAMSDPCKKLAELFESIA 889  
Db 2593 YDSMFD---QVDDIYLRGLSLSTNSLGLTAA-----TTEIILKLAAYFLLL 2638  
QY 890 VDFPPTGPAEIPSLRPEY--PDFMDKPKDTSYISE--RVIGLFRKVKDKAPQASSI 945  
Db 2639 PNPVK-GLNAKTLPOIVPKSVIEPKHRHQEEMVTRISQLKMFGRMRPAEAKSHFLELL 2697  
QY 946 AIF-----TROVAR-RSYDADMEVDGF 966  
Db 2698 STWPTFGVLHYRLKSVENGHQLPEVILTIKSGIQLLOPKSKEVFKERNYDQIVSVESI 2757  
QY 967 EDYI-----DEAFDYKTEYDNKGLNMDYIGIKTEAEILSGGIMKASKT 1010  
Db 2758 RKTAKIVRLVNTMQGETLDIKDEADEISHLIGY-----MFVTGGAEERGST 2808

## RESULT 15

S37972  
low-temperature viability protein LTV1 - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein YKL143w; protein YKL2  
C:Species: Saccharomyces cerevisiae  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 06-Feb-1998  
R:Ramezani Rad, M.; Xu, G.; Kirchhath, L.; Fritz, C.; Keuchel, H.; Hollenberg, C.P.  
submitted to the Protein Sequence Database, March 1994  
A:Reference number: S37953  
A:Accession: S37972  
A:Molecule type: DNA  
A:Residues: 1-463 <RAM>  
A:Cross-references: EMBL:228143; NID:9486244; PID:9486245; MIPS:YKL143w  
R:Experimental source: strain S288C  
R:Abraham, P.R.  
submitted to the EMBL Data Library, August 1993  
A:Reference number: S38996  
A:Accession: S38997  
A:Molecule type: DNA  
A:Residues: 1-134, /ATCVSKRDKGA',146-463 <ABR>

A:Cross-references: EMBL:225464; NID:9396434; PID:9396436  
R:Abraham, P.R.; Mulder, A.; van't Riet, J.; Plant, R.J.; Raue, H.A.  
Yeast 8, 227-238, 1992  
A:Title: Molecular cloning and physical analysis of an 8.2 kb segment of chromosome 2  
A:Reference number: S22274; MUID:92245761; PMID:1574929  
A:Accession: S22275  
A:Molecule type: DNA  
A:Residues: 1-134, /ATCVSKRDKGA',146-313, /LSKTKARRGRRKRAVKKRA',334-363, /SSMVKL' <AB>  
A:Cross-references: GB:S97962  
C:Genetics:  
A:Gene: SGD:LTV1  
A:Cross-references: SGD:S0001626; MIPS:YKL143w  
A:Map position: 11L

## Query Match 2.2%; Score 131; DB 2; Length 463;

Best local similarity 18.0%; Pred. No. 0.52;  
Matches 96; Conservative 75; Mismatches 155; Indels 206; Gaps 24;

QY 530 KISGDFAH-RVASKCGLQYTPSAFQIRYGYGVGVGVPDSSMKLSLRKSMYSKSDNIK 588  
Db 59 KPGRRRAHVGEAALYGINFDDSEYD--YTQHLKPIGLDPENSIFIAKSGNEQKVEKKNIE 116  
QY 589 LDVLGWSKYQPCVNLNQLITLLSTLGVKDEVELEKQKEAVDQDAILHDSLKAQEALELM 648  
Db 117 -DLFIEPKYR-----RDEI-----EKDDALPVPQGMKPE---YLL 149  
QY 649 SPGENTNILKAMLCNGYKPAEPLSMMLQTFRASKLLDLRTRSRIFIPNGRMTMGCLDE 708  
Db 150 HQQDTTDEIR-----GFKPDMNPALREVLEA-----EDEAVVNDVVDVEDISK 194  
QY 709 SRLLE---YGO-----VFQFTGAGHG-----EFSDDLHFPNNSRSTNSNFIKGNVVA 755  
Db 195 KTLQGGDNYGEEKEDDIFAQLLSGEAKDEDEFDE----- 231  
QY 756 KNPCLHPGDIRVLKAVNVRALHNMVDCVFPQKGRPHNECSGSDLDGDIYFVCWDQDM 815  
Db 232 ----- 239  
QY 816 I-----PPROPVQMEYPPAPSIQDHDHVTIEVEEYFTNYIVNDSLGLIANAHVVFADREP 871  
Db 240 VENFEDENYVVKEM-----AQFDNIENLEDLEN-----IDYQA 271  
QY 872 DMAMSDPCKKLAELFSTAVDFPKTGVPAETPSOLRP--KEYPDFMDKPKDTSYISERVIG 929  
Db 272 DVRRFQKDNILEKHNSDDEFSNAGLDSVNPSEEDVGLGELPSIQDK-SKT-----G 322  
QY 930 KLFRKVKDKAPQASSIATFTRD---VARRSY-----DADMEVDGFEDYIDE----- 972  
Db 323 KKKRKSQKKGAMSDVSGFSMSSSAIARTETMTVLDQYDQIINGYENYEELEDEBEQN 382  
QY 973 --AFDY---KTEYDNKGLNMDYIGIKTEAEILSGG---IMKASKTFDRRKA 1017  
Db 383 YQPDFMSAERSDFESMLDDFLDNY-----ELESGRKRLAKKKEIERLKEA 428

Search completed: November 6, 2002, 03:46:51

Job time : 40 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:40:53 ; Search time 15 seconds  
(without alignments)

3080.309 Million cell updates/sec

Title: US-09-782-874-2

Perfect score: 5864

Sequence: 1 MGKTIQVFGPYLLSAEVYK.....RPVLNLSRLAQLSHRLVLK 1114

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	131	2.2	463	1	LTVL_YEAST
2	129	2.2	1733	1	RPBL_YEAST
3	121.5	2.1	2032	1	CTOG_HUMAN
4	118.5	2.0	503	1	DHE3_CHICK
5	117	2.0	688	1	YEQ7_YEAST
6	116.5	2.0	1200	1	MOG5_CABEL
7	116	2.0	950	1	ORP1_HUMAN
8	114.5	2.0	2214	1	PVRL_YEAST
9	113.5	1.9	878	1	SYA_THETN
10	113	1.9	787	1	XPKA_LACPE
11	113	1.9	1748	1	YNR2_YEAST
12	112.5	1.9	988	1	TERT_SCHPO
13	111.5	1.9	891	1	DPO1_HELPY
14	111.5	1.9	1816	1	LMA4_HUMAN
15	111	1.9	648	1	RAD4_SCHPO
16	111	1.9	1014	1	ODOL_YEAST
17	109.5	1.9	642	1	NSG2_ACHCY
18	109.5	1.9	1579	1	SSK2_YEAST
19	109.5	1.9	4085	1	RRPA_CVH22
20	109	1.9	501	1	DHE3_BOVIN
21	109	1.9	909	1	ODOL_BUCAI
22	109	1.9	1170	1	SMC2_YEAST
23	109	1.9	1201	1	RPOC_LISIN
24	109	1.9	1887	1	FAS2_YEAST
25	108.5	1.9	1188	1	YEC5_YEAST
26	108.5	1.9	1835	1	DURL_YEAST
27	108	1.8	734	1	METE_THEMA
28	108	1.8	1607	1	M3K4_HUMAN
29	108	1.8	1916	1	RIF1_YEAST
30	108	1.8	3061	1	POLG_PVYHU
31	107.5	1.8	755	1	P100_HSV7J
32	107.5	1.8	1630	1	MSP1_PLAFK
33	107.5	1.8	1639	1	MSP1_PLAFW
					P34078 saccharomyc
					P04050 saccharomyc
					Q14008 homo sapien
					P00368 gallus gall
					P40050 saccharomyc
					Q09530 caenorhabdi
					Q9bwx6 homo sapien
					P07259 saccharomyc
					Q8raM4 thermoanaer
					Q337f6 lactobacilli
					P33886 saccharomyc
					O13339 schizosacch
					P56105 helicobacte
					Q16363 homo sapien
					P32372 schizosacch
					P20967 saccharomyc
					P34127 achromobact
					P33599 saccharomyc
					Q05002 human coron
					P00366 bos taurus
					P57388 buchera ap
					P38989 saccharomyc
					P7879 listeria in
					P19097 s fatty aci
					P39991 saccharomyc
					P32528 saccharomyc
					Q9x112 thermotoga
					Q9y6r4 homo sapien
					P39539 saccharomyc
					Q02963 p genome po
					P2519 human herpe
					P04932 plasmodium
					P04933 plasmodium

## RESULT 1

LTVL\_YEAST  
ID LTVL\_YEAST STANDARD; PRT; 463 AA.  
AC P34078;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Low-temperature viability protein LTvl.  
GN LTvl OR YKl143W OR YKl2.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID:4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92245761; PubMed=1574929;  
RA Abraham P.R., Mulder A., Van'T Riet J., Planta R.J., Raue H.A.;  
RT "Molecular cloning and physical analysis of an 8.2 kb segment of  
RT chromosome XI of Saccharomyces cerevisiae reveals five tightly linked  
RT genes.";  
RL Yeast 8:227-238(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Rad M.R., Xu G., Kirchraeth L., Fritz C., Keuchel H., Hollenberg C.P.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO S. POMBE SPAC3F10.17.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; Z25464; CAA80955.1; -;  
DR EMBL; Z28143; CAA81984.1; -;  
DR PIR; S37972; S37972.  
DR SGD; S0001626; LTVL.  
FT CONFLICT 135 145 LPVFORGMKP -> ATCVSKRDGKA (IN REF. 1).  
SQ SEQUENCE 463 AA; 53390 MW; F432C239B3B0A68A CRC64;

## Query Match

Best Local Similarity 2.2%; Score 131; DB 1; Length 463;  
Matches 96; Conservative 75; Mismatches 155; Indels 206; Gaps 24;

QY 530 KISGDEAH-RVASCGLOQYTPSAFQIRYGYKGVGVDPDSSMKLSLRKSMKSYEDSNIK 589  
DB 59 KPGRRRAHVEAALYGINFDSEYD--YTQHLKPIGLDPENSIFIAKSGNEQKVEKNIE 116  
QY 589 LDVLGWSKYPCYLNRLQTLTLTLGLVKQKQEAVDQLDLHLSLKQAEALELM 648  
DB 117 -DLFIEPKYR-----RDEI-----EKDDALPVPFORGMKPPE---YLL 149  
QY 649 SPGENTNLLKMLNCGYKPDAPFELSNMQLTFRASKLLDLRLTRTSRIPIPNRTMAGCLDE 708

P19199 commelina y  
Q51561 pseudomonas  
P35194 saccharomyc  
Q9hm08 thermoplas  
Q8ya96 listeria mo  
P06909 mus musculu  
P10844 clostridium  
P32657 saccharomyc  
Q67163 aquifex aeo  
P36336 marek's dis  
Q05783 mycobacteri  
P53935 saccharomyc

## ALIGNMENTS



Db 150 HQDPTDEIR-----GFKPDMNPALEVRLEAL-----EDEAYVNDVVVEDISK 194  
 Qy 709 SRTLE---YGO-----VFVQFTGAGHG-----EFSDDLHPFNNSRSTNSFILKGNVVA 755  
 Db 195 KTOLOGNDYGEEREKEDIFAQLLGSGEAKDEDFEDE-----231  
 Qy 756 KNPCLHPGDIVLKVAVNRALHHMVDVCFVQKGRPHNCEGSLDGLDIYFVCDQDM 815  
 Db 232 -----FDEWDIDN 239  
 Qy 816 I---PPRQVPMYPPAPSTQLDHDVTIEVEYFTNYIVNDSLGLIAHNVVVFADREP 871  
 Db 240 VENFEDENYKEM-----AQFDNIENLEDLEN-----IDYQA 271  
 Qy 872 DMAKSDPCKKLAEILFSAIVDPKTPGPAEIPSQLRP---KEYPDMKPKDPTYSISERVIG 929  
 Db 272 DVRRFQKDNSILEKHSNDEDFSNAGLSDVNPSEEDVLGELPSIQDK-SKT-----G 322  
 Qy 930 KLFKVKDKAPQASSIATETRD-----VARRSV-----DADMEVDGPFEDYIDE-----972  
 Db 323 KKKRSQKKGAMSDVSGFSMSSSAIARTETMTVLDQYDQIINGYENYEELEDEBEQN 382  
 Qy 973 --AFDY---KTEYDNKLGMLMDYIGIKTEAEILSGG---IMKASKTFDRKDA 1017  
 Db 383 YQPFDMGAERSDFESMLDDFLDNY-----ELESGRKLAKDKREIERLKEA 428

RESULT 2  
 APBI\_YEAST STANDARD; PRT; 1733 AA.  
 ID RPBI\_YEAST  
 AC P04050; Q12364; Q92315;  
 DT 01-NOV-1986 (Rel. 03, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE DNA-directed RNA polymerase II largest subunit (EC 2.7.7.6) (B220).  
 GN RPBI OR RPO21 OR RPB220 OR SUB8 OR YDL140C OR D2150.  
 OS Saccharomyces cerevisiae (baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID:4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A364A;  
 RX MEDLINE=85282617; PubMed=3896517;  
 RA Allison L.A., Moyle M., Shales M., Ingles C.J.;  
 RT "Extensive homology among the largest subunits of eukaryotic and  
 RT prokaryotic RNA polymerases";  
 RL Cell 42:599-610(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288c / FY1679;  
 RX MEDLINE=97127826; PubMed=8972577;  
 RA Woelfl S., Haneman V., Saluz H.P.;  
 RT "Analysis of a 26,756 bp segment from the left arm of yeast  
 RT chromosome IV";  
 RL Yeast 12:1549-1554(1996).  
 RN [3]  
 RP SEQUENCE OF 1669-1733 FROM N.A.  
 RC STRAIN-S288c;  
 RX MEDLINE=95377607; PubMed=7649444;  
 RA Cronan J.E. Jr., Wallace J.C.;  
 RT "The gene encoding the biotin-apolipoprotein ligase of Saccharomyces  
 RT cerevisiae";  
 RL FEMS Microbiol. Lett. 130:221-230(1995).  
 CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA}(n).  
 CC -!- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.  
 CC THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE II.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.

CC THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND TRNA GENES.  
 CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; X03128; CAA26904.1; -  
 CC EMBL; X96876; CAA65619.1; -  
 CC EMBL; Z74188; CAA98713.1; -  
 CC EMBL; U27182; AAC49058.1; -  
 CC PIR; A00692; RNBX2L.  
 CC SGD; S0002299; RPO21.  
 CC InterPro: IPR000684; RNA\_pol\_II\_repeat.  
 CC InterPro: IPR000722; RNA\_pol\_A.  
 CC InterPro: IPR002879; RNA\_pol\_A2.  
 CC Pfam; PF00623; RNA\_pol\_A; 1.  
 CC Pfam; PF01854; RNA\_pol\_A2; 1.  
 CC PROSITE; PS00115; RNA\_POL\_II\_REPEAT; 22.  
 CC Transferrase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
 CC DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.  
 CC ZN\_RING 67 83  
 CC C2H2-TYPE (POTENTIAL).  
 CC DOMAIN 1544 1719  
 CC CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
 CC VARIANT 1653 1659  
 CC MISSING (IN STRAIN A364A).  
 CC FT CONFLICT 1514 1514  
 CC A -> V (IN REF. 1).  
 CC FT CONFLICT 1524 1524  
 CC G -> A (IN REF. 1).  
 CC FT CONFLICT 1601 1601  
 CC T -> M (IN REF. 1).  
 CC SQ SEQUENCE 1733 AA; 191610 MW; A45C1360FF99F968 CRC64;  
 CC  
 CC Query Match 2.2%; Score 129; DB 1; Length 1733;  
 CC Best Local Similarity 17.1%; Pred. No. 1.9; Indels 376; Gaps 48;  
 CC Matches 204; Conservative 167; Mismatches 167; Gaps 48;  
 CC  
 CC QY 57 SADKIITLANNR-----LYFGSSYLKAWEMKTDIVQLRAYVDMQDITLNFQGISDDKF 111  
 CC Db 409 SGDRIDLRYSKRAGDIQLQYG-----WKVERHI-----MDNDPVLNRPQSLHRM 453  
 CC QY 112 AVLGSTEVSTQFGIGLKKFFFLSSGSADYKQLQLSYENIW-----QVVLHRYPGQNAQ 164  
 CC Db 454 SMAHRVKVIPY-----STFRLNLSVTSYNADFQDGMNHLVPPQSEETR 498  
 CC QY 165 FLTLTQFGAPR--IYKRLNSCYSFKEK-----PDDQWVTRTTDFPSPWIG 208  
 CC Db 499 AELSQLCAVFLQIVSPQSNKPCMGIVQDTLCGIRKLTLRDTFIELQVLNMLYVVPWDWG 558  
 CC QY 209 L-----SSSLCLQFRGRVRLPNFEESFFHYAERENITLTGTGTFP-----249  
 CC Db 559 VIPTPALIKPKPLWSGKQLSVAIPNGIHLQRFDEGTTLLSPKDNGLIIDGQIIIGVWE 618  
 CC QY 250 ----VSQKSAIVNPVQPEGISIPYKILFKISSIVQHGC-----PGPALNV 292  
 CC Db 619 KKTVGSSNGGLIHVVTRKPGQVCAKLFQGNVQVFNWLLHNGFSGTIGTADGPTMRE 678  
 CC QY 293 YFFRLVDPRRR-----NVACIEHALEKLYIYKECCYD-FVRWLTSEQYDGLYKGRQ 341  
 CC Db 679 IETIEAKKKVLDVTWEAQANLLTAKHGMT-----LRESFEDNVVRFELNEARD-----K 728  
 CC QY 342 PPKSPSTLDDGLVYVRVRLVTPCKYFCPEVNVSNRVLNRYSEDID-NFLRVSFVDEE 400  
 CC Db 729 AGRLAEYNLKD-LNNVKQMYMAGSK-----GSFINIAQMSACVGGQSVSGRIAFGVDR 783  
 CC QY 401 WEKLYSTDLLPKASTGSGVTRNIYERILTLRGKVFIDGKKFEFLAFSSQLRDNVSMVF 460  
 CC Db 784 LPHFSKDDYSPES-----KGFV-----ENSVLR-----806

```
QY 461 ASRFGLTAND--IRAWMGDFSQIKNVAKYARLQCSGSSGRETSLVLRHIEVDPVKVH 518
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 807 ---GLTPQEFPPHMGSGREGSLDTAVKTA-----ETGYIQRRVLVKALEDIMVH 851
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 519 -----GTSYVF---SDGIGKISGDFAH-RVASKCGLOQYPSAQIIRYGGYKGVGV 565
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 852 YDNTRNSLGNVIGFIYGEDGM-----DAHIEKQSLDTIGGSDAAPEKRY-----RVDLL 902
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 566 DPDSKMSLSKSKSYESONIKDLVLGWSKYQPCYNLRNOLITLL-----610
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 903 NTDHTLPSLLESSEILGD-LKLOVLLDEYKOLVDRFLREVFVDGSEANWPLPNVIR 961
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 611 -----STLGVKD-----EVLQKQKEAVDQDLAI 634
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 962 RIIONAQOTFHIDTKPSDLTIKDIVLGKDLQENLLVLRLKNEIIQAQRDAVTLFCCL 1021
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 635 LHSLSKQAEALE-----LMSPGENTNLIKAMLCNGVKPDAEP 671
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1022 LRSRLATRRVQLQYRLTKQAFDWVLSNIEAQFLRSVYHPGEMGVGLAA-----QSIGEP 1075
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 672 FLSMMLQTFRASKLLDLRTRSRIFIPNGRTMMGCLDESRTLEYGQVFQVFTGAGHGFSD 731
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1076 ATQMTLNTFFHAGVASKKVTSG--VPLRLKEILNVAKNMKT-----PSLTVYLEPGH---AA 1126
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 732 DLHPFNRSRSTNSFIKGNVVVAK-----NPLHPGDRLVRLKAVNRALHMHVDCVV 784
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1127 DOEAKLIRSAIEHTLKSVTIASEIYDDPRSTVPIPEDEEIIQ-----LHFSLLDEE 1180
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 785 FPQKGRPHP-----NECSGSDLD--GDI-----YFVCWDQD-----MIPP 818
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1181 AEQSFQDQSPWLLRLLELDRAAMNKDLTMQGVGERIKQTFKNDLFIWSEDNDEKLIIRC 1240
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 819 RQVQPMYPPAPSLQLOHDVTIEVEEYFTNYIYVDSLGIIANAHVYFADR-----869
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1241 RWRPKSLDAETAEAEEDH--MLKXIENITLENITLRGVNIERYVMYKDYRKVPSPTEGY 1298
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 870 --EPDMASDPCKLAELFSTAVDFPKTPVPAEIPSQLRKPKEYPDFMDK-----916
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1299 VKPEPWLETDGVLSEVMT-----VPGIDPTRIYNSFDIMEVLGIEAGRAALY 1349
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 917 -----PKTYSYSERSVIGKLFKRVKDKAPQASSIATFRDVARRSYDADMEVDGDEDY 969
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1350 KEVNVIASDGSYNYRHMALLV---DVMTQGGTLSTVRHFNRSNTGALMCRSPFEET 1405
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 970 IDEAPDYKTEYDNKLGNDYGIKTAELISGGIMKASKTFDRRKDAEAI 1020
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1406 VELLFEAGAS-----AELDDCRGV-SENVILGQAPICGTGAFDVMIDEESL 1450
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

RESULT 3
CTOG_HUMAN STANDARD; PRT; 2032 AA.
ID CTOG_HUMAN Q14008; Q14668;
AC Q14008; Q14668;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE CH-TOG protein (Colonic and hepatic tumor over-expressed protein).
OS KIAA0097.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain tumor;
RX MEDLINE=96128167; Pubmed=8536682;
RA Charrasse S., Mazel M., Taviaux S., Berta P., Chow T., Larroque C.;
RT "Characterization of the cDNA and pattern of expression of a new gene
over-expressed in human hepatomas and colonic tumors."
RL Eur. J. Biochem. 234:406-413(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
```

```
RX MEDLINE=95308325; Pubmed=7788527;
RA Nagase T., Miyajima N., Tanaka A., Sazuka T., Seki N., Sato S.,
RA Tabata S., Ishikawa K.-I., Kawabayasi Y., Kotani H., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. III.
The coding sequences of 40 new genes (KIAA0081-KIAA0120) deduced by
analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 2:37-43(1995).
CC -!- TISSUE SPECIFICITY: OVER-EXPRESSED IN HEPATOMAS AND COLONIC
TUMORS. ALSO EXPRESSED IN SKELETAL MUSCLE, BRAIN, HEART, PLACENTA,
LUNG, LIVER, KIDNEY AND PANCREAS.
CC -!- SIMILARITY: CONTAINS 9 HEAT REPEATS.
CC -!- SIMILARITY: TO C.ELEGANS F22B5.7 AND F22B5.8 AND SOME, TO YEAST
STU2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; X92474; CAA63212.1; -
DR EMBL; D43948; BAA07892.2; ALT_INIT.
DR InterPro: IPR000357; HEAT_repeat.
DR PROSITE; PS50077; HEAT_REPEAT; 1.
KW Repeat.
FT REPEAT 160 197 HEAT 1.
FT REPEAT 357 394 HEAT 2.
FT REPEAT 435 472 HEAT 3.
FT REPEAT 751 788 HEAT 4.
FT REPEAT 937 974 HEAT 5.
FT REPEAT 1014 1051 HEAT 6.
FT REPEAT 1285 1322 HEAT 7.
FT REPEAT 1324 1357 HEAT 8.
FT REPEAT 1362 1399 HEAT 9.
FT CONFLICT 1564 1623 MISSING (IN REF. 1).
FT CONFLICT 1814 1814 E -> A (IN REF. 1).
FT CONFLICT 1822 1822 E -> A (IN REF. 1).
SQ SEQUENCE 2032 AA; 225507 MW; 0256603047FA45EF CRC64;

Query Match 2.1%; Score 121.5; DB 1; Length 2032;
Best Local Similarity 22.4%; Pred. No. 8.1;
Matches 110; Conservative 66; Mismatches 146; Indels 169; Gaps 26;

QY 276 ISSLVHGGCPGPNLVYFRLVDPRRNVACIEHALEKLYIKCCYDPVRMLTQYDQ 335
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 565 VALIAGKGNFSKTSAQVLDGLVD-KIGDKCGNNKAEMATAEACMLP--WTAEOVVS 721
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

QY 336 YLKGROPPKSPSTILD-----DGL-----VYVRVLVTPCKV 367
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 722 MAFSQKNPKNQSETLNLWSNAIKFEGFSGNLNVKAFISNVKTALATNPVTRTAITLLGV 781
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

QY 368 --YFCGPEVNVSNRNLNYSIEDINFLRVSVFVDEWEKL-----YST----407
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 782 MYLVGFS-----LWMFEDEKPAL-LSQIDAEKMGQSPAPTRGISNKHSTSGTD 833
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

QY 408 -----DLLPKASTGSGVTRNIYERILSTLRKGFVIGDKKFEFLAFSSSQ 451
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 834 EGEDGDEPDGSDNDVVDLLP-----RTEISDKITSELVS--KIGDKNWKI-----876
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

QY 452 LRNSVWVFASRPGLTANDIRANWGDPSQIK-NVAKYAARLGOSFGSS-----RETLVL 505
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 877 -----RKEGL--DEVAGIINDAKFTQPNIGELPTALKGRLNDNSKILVQQTINIL 924
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

QY 506 RH-EIEVIPDKVHGTSYVFSGIGKIS--GDFAHRVASKCGLOQYTPSAQIIRYGGYKGV 562
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 925 QQLAVAMGNPKQH-----VKNLGIPIITVLGDSKNVY--RAAALATVNAAEQTGMKEWL 978
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

QY 563 VGVDPDSSKMSLSKSKSYESDNKLDVLGWSKYQPCYNLRNOLITLLSTLGVKDEVLEQ 622
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 979 EGEDLSEEL-----KKENPFLRQELLGW-----LAELKPLTRST--PTDLIL--1018
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
```



QY 674 SMLQTFRASK-----LLDLRSTRIFIPNGRTMMGCLDESKTLEYGVQV 717  
Db 313 LMIRSCRIADKREVMVNDLANK-----IPD-----YCVDSKLLFEYCEV 354

## RESULT 6

MOG5\_CAEEL STANDARD; PRT; 1200 AA.  
AC Q09530;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Probable pre-mRNA splicing factor ATP-dependent RNA helicase mog-5  
DE (sex determination protein mog-5) (Masculinization of germ line  
DE protein 5).  
GN MOG-5 OR EED8.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
[1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=20202625; PubMed=10737793;  
RA Puoti A., Kimble J.;  
RT "The hermaphrodite sperm/oocyte switch requires the Caenorhabditis  
RT elegans homologs of PRP2 and PRP22".  
RL Proc. Natl. Acad. Sci. U.S.A. 97:3276-3281(2000).  
[2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=Bristol N2;  
RA Chisoe S.;  
RL Submitted (JUL-1995), to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: COULD ACT LATE IN THE SPLICING OF PRE-MRNA AND MEDIATE  
CC THE RELEASE OF THE SPLICED MRNA FROM SPICEOSOMES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAH  
CC SUBFAMILY. DD8/PRP22 ORTHOLOG.  
CC -1- SIMILARITY: CONTAINS 1 SL MOTIF DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL; AF286899; AAC01332.1; -  
DR EMBL; U23484; AAC4765.1; -  
DR HSSP; P05055; ISRO.  
DR WormPep; EED8.5; CE01889.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR002464; DEAH\_box.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR003029; SL.  
DR Pfam; PF00271; helicase\_C; 1.  
DR Pfam; PF00575; SL; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR SMART; SM00490; HELIC; 1.  
DR SMART; SM00316; SL; 1.  
DR PROSITE; PS00690; DEAH\_ATP\_HELICASE; 1.  
DR PROSITE; PS0126; SL; 1.  
KW mRNA processing; mRNA splicing; Helicase; ATP-binding;  
KW Nuclear protein; Developmental protein.  
FT DOMAIN 176 185 POLY-ARG.  
FT DOMAIN 234 304 SL MOTIF.  
FT NP\_BIND 563 570 ATP (POTENTIAL).  
FT SITE 660 663 DEAH BOX.  
SQ SEQUENCE 1200 AA; 135759 MW; 023C46E494D7F44F CRC64;

Query Match 2.08; Score 116.5; DB 1; Length 1200;  
Best Local Similarity 18.88; Pred. No. 7.9;  
Matches 166; Conservative 137; Mismatches 331; Indels 247; Gaps 44;

QY 328 WLTEQYDGLKGRQPKSPSIITLDDGLVYVRRVLVTP-----CKYVFCG-----PEVN 375  
Db 340 WNPESAGVGQG-----PSTSIGGKARVR--ISTPERWELRQMGAGVLTATDMPDFD 391  
QY 376 VSNRVLNYSIEDID-NFLRVSVFVDEWEKLY-----STDLLP-----KASTGSGVRTNIYE 425  
Db 392 EEMGVLNRYDDESDGEDIIEILVEDPDLRGYKGGAIEPVKVVKNPDGSLAQAALMQ 451  
QY 426 RILSTLRKGFVIGDKKFEFL-----AFSS--SQLRDNSSVMMFASRPGILTANDIRAWGDFSQ 480  
Db 452 GALSKEKRTKTOAQERDMDTQKGFSSNARILD-----PMSGNQSTAWSADESK 501  
QY 481 IKN-----YAKYARALCQSGSSREILSVL---RHELEIVPDVKVHGT 520  
Db 502 DRNNKMKEMPEWLKHVYTAGKATYGRRTNLSMVEQRESLPIFALKKNLMEAMIDNQI--L 559  
QY 521 SVFSDGIGKISGDFAHRVASKGQYTPSAFQIRYGYKGVVG-VDPDSSMKLSLRKSM 579  
Db 560 VVVGEGSGKTT-----QMTQYALEAGL-GRRGKIGCTQPRVAAMSAVKRV 605  
QY 580 SKYESDNIKDLVGSKYQPCYLNQLITLSTLGVKDEVLEQKQKEAVDQDLALHDSL 639  
Db 606 ABEYGCGLGTDVGYTIRFEDCTSQDTIIKYMT-----DGML---LRECLIDPDLSGYSLI 657  
QY 640 KAAQEALELMSPGENTNLIKAMLCNGYKPDAPFLSNMLQTFRASKLLDLRSTRIFPNG 699  
Db 658 MLDEAHERTI---HTDVLGLKAAARKRPKLIITSLDVSKEFSEFLEAPITFPG 714  
QY 700 RTMMGCLDESRTLEYGQVEVQFTGAGHGEFSDDLHPFNNSRSTNFIKGNVYVAKNCP 759  
Db 715 RTF-----PVEILYT-----REPESDYLEAAHITVMQ--- 741  
QY 760 LH-----PGDIRVL---KAVNVRALHVMVDCVVFQPKGRPHNCSGSDLDGDIYFVCW 811  
Db 742 IHLTEPPGDLVFLGQEEID-----TSCEVLIERK-----SMGPDVPELI----- 783  
QY 812 DQDMIP-----PRQVQPMPEYPAPSIQDHDVTIEEVEEFTNYIVNDSLGIIANAHVVF 866  
Db 784 ---ILPVGALPSEMOTRIFDPAPAGK--RKVVIA-----TN-IAETSLTIDGIFYV- 830  
QY 867 ADREPDMASDCKKLAEILFSTAVDPKGTGVAEPISQLRPEYDFMDKPKDTSVISER 926  
Db 831 -----DGFVKQKIYN-----PRSGMDSLVTVPISQAAAKQSRGRGTG----- 870  
QY 927 VIGKLPFRKVKAPQASSIATFRDVARSYDA---DMEVDGFEYIDFAFYKTEYDN- 982  
Db 871 -PGKCYRLYTERAFRDEMPLTPVPEIQRTNLAETLLQLKAMGINNLLIDPFDMDAPLDSM 929  
QY 983 -----KLGNLMDYGIKITEAIELSGGIMKASKTFDRKRD---AEA 1019  
Db 930 ITALNTLHTLSALDGDGLTLKLRNNAEPFLEP-----SUSKLLIMSVDLGCSEE 979  
QY 1020 ISVAVRALRKEARWPKRRNDIDMLPKASAWYHVTY--HPTYWCYNOGLKRAHFISFP 1077  
Db 980 VLTIVAML--NYQNIFYRPKEQDHADQKAKFHOPEGDHLLTLLAVYNSW--KNHHSQP 1035  
QY 1078 WCYVDQLIQIKDKKARNRPVNLSSLAQL-----SHRLVL 1113  
Db 1036 WC-FENFIQVRSMK-----RAQDIRKOLLGIMDRHKLIM 1068

RESULT 7  
ORP1\_HUMAN STANDARD; PRT; 950 AA.  
ID ORP1\_HUMAN Q9BW87; Q9B2F5;  
AC Q9BW87; Q9BW87 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Oxysterol binding protein-related protein 1 (OSBP-related protein 1)  
DE (ORP-1).  
GN OSBP1 OR ORP1 OR OSBP8.  
OS Homo sapiens (Human).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
NCBI\_Taxid=9606;  
[1]  
SEQUENCE FROM N.A. (ISOFORMS A AND B).  
JAWORSKI C.J., MOREIRA E., LI A., LEE R., RODRIGUEZ I.R.;  
"A family of 12 human genes containing oxysterol-binding domains.";  
Genomics 78:185-196(2001).  
[2]  
SEQUENCE FROM N.A. (ISOFORM A).  
MEDLINE=21264638; PubMed=11279184;  
Xu Y., Liu Y., Ridgway N.D., McMaster C.R.;  
"Novel members of the human oxysterol-binding protein family bind  
phospholipids and regulate vesicle transport.";  
J. Biol. Chem. 276:18407-18414(2001).  
[3]  
SEQUENCE FROM N.A. (ISOFORM C).  
MEDLINE=21376257; PubMed=11483621;  
Lehto M., Lahtinen S., Chinetti G., Johansson M., Ehrenholm C.,  
Staelen B., Ikonen E., Oikarinen V.M.;  
"The OSBP-related protein family in humans.";  
J. Lipid Res. 42:1203-1213(2001).  
[4]  
SEQUENCE OF 514-790 FROM N.A. (ISOFORM A).  
TISSUE=Embryo;  
Isogai T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,  
Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
Tanase T., Nomura Y., Togiya S., Komai F., Hara K., Takeuchi K.,  
Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
Wakamatsu A., Nakamura Y., Nagahari K., Masuko Y., Oshima A.;  
"NEO human cDNA sequencing project.";  
Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.  
CC -!- FUNCTION: Binds phospholipids; exhibits strong binding to  
phosphatidic acid and weak binding to phosphatidylinositol 3-  
phosphate (By similarity).  
CC -!- ALTERNATIVE PRODUCTS: 3 isoforms; A/OSBPL1A/OSBP8S,  
B/OSBPL1B/OSBP8L (shown here), and C; are produced by alternative  
splicing.  
CC -!- SIMILARITY: BELONGS TO THE OSBP FAMILY.  
CC -!- SIMILARITY: CONTAINS 3 ANK REPEATS.  
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to licenses@isb-sib.ch).  
-----  
DR EMBL; AF392449; AAL40562.1; -  
DR EMBL; AF392450; AAL40563.1; -  
DR EMBL; AF274714; AAK15154.1; -  
DR EMBL; AF323726; AAG53407.2; -  
DR EMBL; AK001079; BA91496.1; -  
DR Genew; HGNC:16398; OSBPL1A.  
DR MIM; 606730; -  
DR InterPro; IPR002110; ANK.  
DR InterPro; IPR000548; Oxysterol\_BP.  
DR InterPro; IPR001849; PH.  
DR Pfam; PF00169; PH; 1.  
DR Pfam; PF01237; Oxysterol\_BP; 1.  
DR SMART; SM00248; ANK; 3.  
DR SMART; SM00233; PH; 1.  
DR PROSITE; PS50088; ANK\_REPEAT; 3.  
DR PROSITE; PS50297; ANK\_REP\_REGION; 1.  
DR PROSITE; PS01033; OSBP\_1.  
DR PROSITE; PSS00003; PH\_DOMAIN; 1.  
DR KW Lipid transport; Transport; ANK repeat; Repeat; Alternative splicing.  
FT REPEAT 47 76 ANK 1.  
FT REPEAT 80 109 ANK 2.  
FT

OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 28583 / FL100;  
 RX MEDLINE=89378778; PubMed=2570735;  
 RA Souciet J.-L., Nagy M., le Gouar M., Lacroite F., Potier S.;  
 RT "Organization of the yeast URA2 gene: identification of a defective  
 RT dihydroorotase-like domain in the multifunctional carbamoylphosphate  
 RT synthetase-aspartate transcarbamylase complex.";  
 RL Gene 79:59-70(1989).  
 RN [2]  
 RP SEQUENCE OF 1-510 FROM N.A.  
 RX MEDLINE=87286375; PubMed=3039294;  
 RA Souciet J.-L., Potier S., Hubert J.-C., Lacroite F.;  
 RT "Nucleotide sequence of the pyrimidine specific carbamoyl phosphate  
 RT synthetase, a part of the yeast multifunctional protein encoded by  
 RT the URA2 gene.";  
 RL Mol. Gen. Genet. 207:314-319(1987).  
 RN [3]  
 RP SEQUENCE OF 1268-2214 FROM N.A.  
 RX MEDLINE=89255278; PubMed=2498313;  
 RA Nagy M., le Gouar M., Potier S., Souciet J.-L., Herve G.;  
 RT "The primary structure of the aspartate transcarbamylase region of  
 RT the URA2 gene product in Saccharomyces cerevisiae. Features involved  
 RT in activity and nuclear localization.";  
 RL J. Biol. Chem. 264:8366-8374(1989).  
 RN [4]  
 RP SEQUENCE OF 1-276 FROM N.A.  
 RX STRAIN=S288c / FY1679;  
 RC MEDLINE=97103775; PubMed=8948101;  
 RA Cziepluch C., Kordes E., Pujol A., Jauniaux J.-C.;  
 RT "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X  
 RT reveals 19 open reading frames including URA2 [5' end], TRK1, PBS2,  
 RT SPT10, GCD14, RPE1, PHO86, NCA3, ASF1, CCT7, Gdf3, two tRNA genes,  
 RT three remnant delta elements and a Ty4 transposon.";  
 RL Yeast 12:1471-1474(1996).  
 RN [5]  
 RP SEQUENCE OF 175-2214 FROM N.A.  
 RX STRAIN=S288c / FY1679;  
 RC MEDLINE=96408771; PubMed=8913765;  
 RA Katsoulou C., Tzeremia M., Tavernarakis N., Alexandraki D.;  
 RT "Sequence analysis of a 40.7 kb segment from the left arm of yeast  
 RT chromosome X reveals 14 known genes and 13 new open reading frames  
 RT including homologues of genes clustered on the right arm of  
 RT chromosome XI.";  
 RL Yeast 12:787-797(1996).  
 CC -1- FUNCTION: THIS PROTEIN IS A "FUSION" PROTEIN ENCODING THREE  
 CC ENZYMIC ACTIVITIES OF THE PYRIMIDINE PATHWAY (GATASE, CPSASE,  
 CC AND ATCASE).  
 CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +  
 CC phosphate + glutamate + carbamoyl phosphate.  
 CC -1- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-aspartate = phosphate  
 CC + N-carbamoyl-L-aspartate.  
 CC -1- PATHWAY: Pyrimidine biosynthesis; first step.  
 CC -1- PATHWAY: Pyrimidine biosynthesis; second step.  
 CC -1- MISCELLANEOUS: GATASE (GLUTAMINE AMIDOTRANSFERASE) AND CPSASE  
 CC (CARBAMOYL PHOSPHATE SYNTHASE) FORM TOGETHER THE  
 CC GLUTAMINE-DEPENDENT CPSASE (GD-CPSASE) (EC 6.3.5.5).  
 CC MISCELLANEOUS: IN EUKARYOTES EC 6.3.5.5 IS SYNTHESIZED BY TWO  
 CC PATHWAY-SPECIFIC (ARGININE AND PYRIMIDINE) UNDER SEPARATE CONTROL.  
 CC -1- SIMILARITY: THE CPSASE DOMAIN IS SIMILAR TO OTHER CPSASES.  
 CC -1- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE DHOASE FAMILY.  
 CC DEFECTIVE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M27174; AAA68280.1; ALT\_INIT.  
 DR EMBL; X05553; CAA29068.1; -.  
 DR EMBL; J04711; AAA35198.1; -.  
 DR EMBL; Z49405; CAA89425.1; -.  
 DR EMBL; X87371; CAA60825.1; -.  
 DR EMBL; D28139; BAA05680.1; -.  
 DR PIR; S05767; Q2BYU2.  
 DR HSP; P00968; IAGX.  
 DR SGD; S0003666; URA2.  
 DR InterPro; IPR002029; Asp/Orn\_Cotranf.  
 DR InterPro; IPR002082; Asp\_carbmltransf.  
 DR InterPro; IPR001317; CPS\_Gatase.  
 DR InterPro; IPR000901; CPSase.  
 DR InterPro; IPR002474; CPSase\_sm\_chain.  
 DR InterPro; IPR002195; Dihydroorotase.  
 DR InterPro; IPR000991; GATase\_1.  
 DR InterPro; IPR004362; MGS\_like.  
 DR Pfam; PF00117; GATase; 1.  
 DR Pfam; PF00185; OTCace; 1.  
 DR Pfam; PF00289; CPSase\_L\_chain; 2.  
 DR Pfam; PF00744; Dihydroorotase; 1.  
 DR Pfam; PF00988; CPSase\_sm\_chain; 1.  
 DR Pfam; PF02142; MGS; 1.  
 DR Pfam; PF02729; OTCace\_N; 1.  
 DR Pfam; PF02786; CPSase\_L\_D2; 3.  
 DR Pfam; PF02787; CPSase\_L\_D3; 1.  
 DR PRINTS; PRO0100; AOTCASE.  
 DR PRINTS; PRO0098; CPSASE.  
 DR PRINTS; PRO0099; CPSGATASE.  
 DR PRINTS; PRO0096; GATASE.  
 DR TIGRfams; TIGR00670; asp\_carb\_tr; 1.  
 DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.  
 DR PROSITE; PS00442; GATASE\_TYPE\_1; 1.  
 DR PROSITE; PS00866; CPSASE\_1; 2.  
 DR PROSITE; PS00867; CPSASE\_2; 2.  
 KW Pyrimidine biosynthesis; Ligase; Transferase; Multifunctional enzyme.  
 FT DOMAIN 1 400  
 FT DOMAIN 401 440  
 FT DOMAIN 441 1482  
 FT DOMAIN 1483 1821  
 FT DOMAIN 1493 1821  
 FT DOMAIN 1822 1909  
 FT DOMAIN 1910 2214  
 FT ACT\_SITE 302 302  
 FT ACT\_SITE 386 386  
 FT ACT\_SITE 388 388  
 FT CONFLICT 86 86  
 FT CONFLICT 123 123  
 FT CONFLICT 250 257  
 FT CONFLICT 270 270  
 FT CONFLICT 313 314  
 FT CONFLICT 372 373  
 FT CONFLICT 394 402  
 FT CONFLICT 431 433  
 FT CONFLICT 482 482  
 FT CONFLICT 485 485  
 FT CONFLICT 492 492  
 FT CONFLICT 501 510  
 FT CONFLICT 552 852  
 FT CONFLICT 852 852  
 FT CONFLICT 1411 1412  
 FT CONFLICT 1482 1482  
 FT CONFLICT 1588 1588  
 FT CONFLICT 1592 1592  
 FT CONFLICT 1595 1595  
 FT CONFLICT 1937 1937  
 FT CONFLICT 1997 1997  
 FT CONFLICT 2039 2039  
 FT CONFLICT 2158 2165  
 FT CONFLICT 2214 AA; 245124 MW; 4CS58304DRECD21 CRC64;  
 SQ SEQUENCE

Query Match

Best Local Similarity 2.0%; Score 114.5; DB 1; Length 2214;

Matches 153; Conservative 119; Mismatches 267; Indels 273; Gaps 43;

QY	384	YSIEDIDNF-----LRVSFVDEWEKLYSTDJ-----LPKASTGSGVRINIYERILST	430
Dd	347	FAYVDVTLTSGWKPLFVNANDDSNEGIVHSELPSYFSQVHPESPFG-----PRDTEF	398
QY	431	LKRGFVGDKKEEF-----LAFSSQLRDNSVMFASRPGLTANDIRA-WMGDFSQIKN	483
Dd	399	LFDVFIAQVKKEFYQTUULKPAFFPGCLLEDN-----VRAHPRIEAKKVLVLCGGLS----	450
QY	484	VAKYAARLGQSGSREITSLVLRHE-----IEVIPDVKHGTSYFSDGI--GKISGDFAH	537
Dd	451	----IQAGEFDYSGSQAIKALKEBGIITILINPIATIQTKSLADKVFPVPVTAEFVR	506
QY	538	RVASKCGLQYTSPAQIRYGXGVG-VGVDPDSSMKLSRKSMSKYESBDNIKLDVLGWSK	596
Dd	507	KVI-----LHERDAIYVTGGGTALSUGI---AMK-----DEFALGVKV-----	544
QY	597	YOFCYLNRQLITLLST-----LGVKDEVLEQOK- EAVDQLDALHLHDSLKAQEALEIMS	649
Dd	545	----LGTPIDIIITDERELFSNAIDEINEKCAKSAANSVD-----EALAANK	589
QY	650	PCENTMILKAM-----LNCGYKPDAPPEPLSMWLOTFRASKLLDLTRSRIFIPNGRTMMG	704
Dd	590	ETGFPVIVAAVALGGLSGGFANNEKEIVLCNVAFSSP-----	639
QY	705	CIDESRTLEYGVQVFTGAGHEFSDDL--HPENN--SRSTNSGNFILKG-----NYVVA	755
Dd	630	-----QVLVEKSMGWKEYEYEVWRDAFDCITVCNMENFDPGLGHTGDSIWVA	678
QY	756	KNPCLHPGDIVULKAVNVRALHHM-----VDCVVPOKGRPHPNCS	798
Dd	679	PSOTLSDEYNMLRTTAVNIHRLGVWGECNIQYALNPVSKDYCI--EVNARLSRSAL	736
QY	799	GSDLDGDIIYVCWDQMIPROGVQWMEYPAPPISOQLDHDTIEVEEYFTIYVNDSLGI	858
Dd	737	ASKATG-----YPLAYTAA--KLGLNIPLINEVKNSVTK-----	767
QY	859	IANAHVVFADREDPMAMSDFCKLAELFSIADVDFK-----TGVPABIPOLAPKEYPD	912
Dd	768	----STCACFESL---DYC-----VVKMPRWDLKXFTRYSTELSUSM--KSYGE	808
QY	913	FMDPKDPTSIYSERVIGKLFPRVKOKAPOASSIAFTTRDVARRSVADAMEVDGFEY---	969
Dd	809	VMS-----IGTFEEAQIAIRSTEYANLGFNETDLDDIDDYELNNPDMRVF	856
QY	970	IDAEFDYK-----TEYD-----NKLGIMDY-----YGIKTBAEILSGGIMKAS	1008
Dd	857	AIANAFAKKGSYDKVWEMETRICKWLNLKHLDOFAEKISFGTKEELPSL---VLURQA	913
QY	1009	KT--FDRKDA---EAISVAVRALRKE--ARAWFKRRNDIDDMPLPKASAWHYTY---H	1057
Dd	914	KQLGDFDRQIARFLDSNEVAIRLRKEYGITFVQKDITVAAREFFPAYTNLYLMYINADSH	973
QY	1058	PTYW-----GCYNOGLKRAHFISPPWC	1079
Dd	974	DLSPDDHGVMVLGSGVYRIGSS-----VEFDWC	1001

## RESULT 9

SYNTHETIC	SYNTHETIC	STANDARD	PRT	878 AA.
AC	SYNTHETIC			
ID	Q8RAH4			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Alanyl-tRNA synthetase (EC 6.1.1.) (Alanine--tRNA ligase) (AlaRS)			
DE	Alanyl-tRNA synthetase (EC 6.1.1.) (Alanine--tRNA ligase) (AlaRS)			
GN	ALAS OR TPE1248			
OS	Thermoanaerobacter tengcongensis.			
OC	Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;			
OC	Thermoanaerobacteriaceae; Thermoanaerobacter.			
CC	NCBI_TaxID=119072;			
RN	[1]			
RP	SEQUENCE FROM N.A.			





```
RA Obermaier B., Piravandi E., Rinke M., Domdey H.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO A.NIDULANS BLME AND MOUSE TSG24.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z71448; CAA96060.1; -
DR SGD; S00051116; YNL172W.
DR InterPro; IPR002015; APC.proteasome.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 1378 1398
SQ SEQUENCE 1748 AA; 196142 MW; 34BC9212020BB285 CRC64;

Query Match 1.9%; Score 113; DB 1; Length 1748;
Best Local Similarity 19.4%; Pred. No. 24;
Matches 93; Conservative 79; Mismatches 157; Indels 150; Gaps 22;

QY 256 LVPNVQPEGISPKILFKISSIVQGCIPGALNVYFFRLVDPDRRVACIEHALEKL 315
DQ 428 LLPDVLDD-----SIPFKI-----YGNSPQNMIRLENLKLKPSRIQAMYIHELLESC 474
QY 316 YVI-----KECCYDVPVRLTEQYDGLKGRQPK--SPSITLDDGLVYVRRVLVT 363
DQ 475 LILVSEQNKEEKACLYDP-----FVKITSPSKNISELTKQNSLPSLQKLPFY 524
QY 364 P-----CKVYFCGPEVNS-----NRVLNVSEIDNPLRV-----SFV-- 397
DQ 525 PETSFTKLCFCAVYKITSFPAFTSFIFLWQSAVILLSRANDDVVGKMKHEDAFSLVLS 584
QY 398 -----DEENEKLYSTDLLPKASTGSGVRNIVYERI---LSTLRKGF----- 435
DQ 585 LLILPISSSAQEQYQYKEIYERDLFQHLQDSEITSVLPRIVIGLHLIREYSLNVLC 644
QY 436 -----VIGDKKFEFLARSSQLRNSVWFASRGLTANDIRAWMGDFSIKNVAKYAA 490
DQ 645 RNEHALLG-----QFLRFATAAMGPDFLWQSYVFKMDSE----- 679
QY 491 LQSGFGSSRETLSLRHEIEVIPDKVHGTSYVFESDGKIGK-----DFARRVA--S 541
DQ 680 -SKFLHPREQNSIFFPLDEPSSI-----TKSLYS--ITENSSITPLCFFTSFSLVATDT 732
QY 542 KGLQYTPSAFQIRYGYKGVGVVD--PDSSMKLSLRKSKYESDNKLDVLGWSKYQP 599
DQ 733 QVELRITPRSPFKI-LGLYELVHSPNLPDYVLGI---LSSFKVDKDELQTYPLG----- 782
QY 600 CYLNRQLITLSTLGVKDEVLEQKQKAVDQDLDAILHDSLKQAFAELMSGENTNLK 658
DQ 783 -----ILVFLQNIL-----KILEDKLSVRONLELDRLADLRQCSAINTSRSDSEKVIK 832

RESULT 12
TERM_SCHPO
ID TERM_SCHPO STANDARD; PRT; 988 AA.
AC O13339; O13338;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Telomerase reverse transcriptase (EC 2.7.7.-) (Telomerase catalytic
DE subunit).
GN TRP1 OR SPBC29A3.14C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN-972;
RX MEDLINE-97400623; PubMed-9252327;
RA Nakamura T.M., Morin G.B., Chapman K.B., Weinrich S.L., Andrews W.H.,
RT Lingner J., Harley C.B., Cech T.R.;
RL "Telomerase catalytic subunit homologs from fission yeast and human.";
RN Science 277:955-959(1997).
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RX MEDLINE-21848401; PubMed-11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Pearson D., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Woestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakowski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- FUNCTION: TELOMERASE IS A RIBONUCLEOPROTEIN ENZYME ESSENTIAL FOR
CC THE REPLICATION OF CHROMOSOME TERMINI IN MOST EUKARYOTES. IT
CC ELONGATES TELOMERES. IT IS A REVERSE TRANSCRIPTASE THAT ADDS
CC SIMPLE SEQUENCE REPEATS TO CHROMOSOME ENDS BY COPYING A TEMPLATE
CC SEQUENCE WITHIN THE RNA COMPONENT OF THE ENZYME.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: TWO ISOFORMS ARE PRODUCED BY ALTERNATIVE
CC SPLICING.
CC -!- MISCELLANEOUS: DELETION CAUSES TELOMERE SHORTENING AND SENESCENCE.
CC -!- SIMILARITY: BELONGS TO THE REVERSE TRANSCRIPTASE FAMILY.
CC -!- TELOMERASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF015783; AAC49802.1; -
DR EMBL; AF015783; AAC49803.1; -
DR EMBL; AL022299; CAA18391.1; -
DR InterPro; IPR000477; RVTse
DR InterPro; IPR003545; Telomerase_RT.
DR Pfam; PF00078; rvt; 1.
DR PRINTS; PR01365; TELOMERASERT.
KW Transferase; RNA-directed DNA polymerase; Telomere; Nuclear protein;
KW DNA-binding; Alternative splicing.
FT VARSPPLIC 524 524 K -> KQ (IN SECOND ISOFORM).
SQ SEQUENCE 988 AA; 116328 MW; AB2DC7030228F443 CRC64;

Query Match 1.9%; Score 112.5; DB 1; Length 988;
Best Local Similarity 18.2%; Pred. No. 11;
Matches 128; Conservative 109; Mismatches 258; Indels 207; Gaps 30;

QY 134 LSSGSADYKQLQSYENIWQVWLHRPYGQNAQFLLIQFGAPRIYKRLNESCYSFFKETPD 193
```

```
Db 130 LVSTFPNVLISLESKNQ-----LLEIGSDAMHVLKSG--SIFEALPN 174
QY 194 DQWRTTDFP-----PSWIGLSSLCLOFRGV 221
Db 175 DNYLGISGLPKNNVFEVTSKKRRTIETITONKSARKEVSNISIRSFIFRSS 234
QY 222 RLPNEEFSEFFHYAERENITQTGTFEVSQKALPVNQPPEGISIPYKILFISSLV- 280
Db 235 YKFKQDLXFNLSICDRNTVHMLOWLFPQFGLINAFQ-----VKLHKVPLVS 286
QY 281 QHGCTPGPALNVPFRVNDPRRRNVACIEHALEKLYIYKECCYDPVRLWLTBOYDYGKGR 340
Db 287 QSTVVPKRLKYV--PLIEQAKRLERI--SLSKYV-----NHYCPYIDTH 328
QY 341 QPKPSPTLDDGLV--VYRRVLVTPCKVFCGRPVNYSN--RVLNYSIEDTDLRLVSEV 397
Db 329 DDEKILSYSLAPNQVAFELRSILV--RVF---FKLWGNQRIEILKDLTFLKLSRY 382
QY 398 DE-----EWEKLYSTDLLPKASTGSGVRTNIYERILSTLRKGFVIGDKKF 442
Db 383 ESFSLHYLMSNIKISEIEWLGLGRSNAMKCLSDPEKRRQIFAEBFYWLYNSFIPILOS 442
QY 443 EFLAFSSSOLRDSV-----WMPASRPGLTANDIRAW----- 474
Db 443 FYITESDLRNRVYFRKDIWKLCLRPFTITSMKMEAFKINENNVRMDTKTLPPIAVI 502
QY 475 -----MGDFSQIKNAK-YAARLGOS---FGSSRETL-----SVLRH-----EIEVIP-DVK 516
Db 503 RLLPKKNTFRITNLRKRLIKMGSKMVLSTNQTLPVASILKHLNESSGIPFNLE 562
QY 517 VHGTSYVSDGKIGKSGFAHRVASKGLQVTPSAFQIRYGGYGVGVGVDPSMKLSLR 576
Db 563 VYKMLLTKFKDLK-----HRM-----FGKKYFVRIDIKSY----- 595
QY 577 KMSKYESDNIKLD---VLGWSKYQPCYNRLQTLITLTLGVXDEVLEQKQKAVDQD 632
Db 596 -----DRIKQDLMFRLVKKLKDPEVIRKVIATHT---SDRATKNVSEAFSYFD 644
QY 633 AILHDSLKAQEALEMLSPGENTNLIKAMLGKYPDAPFFLSMMLQTFRASKLIDL--- 689
Db 645 MY-----PFEKVQLLSMKTSITLFDVFDVYWKSSSEIF--KMLKEHLSGHIVKIGNSQ 697
QY 690 TRSRIFIPNGRT-----MMGCLDE--SRTLEYGOVEVQ 720
Db 698 YLQKVGFQGSILSFLCHFYMEDLIDEVLSFTKKKGSVLLR 739

RESULT 13
DPO1_HELPY
AC P56105;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA polymerase I (EC 2.7.7.7) (POL I).
GN POLA OR HP1470.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26595 / ATCC 700392;
RX MEDLINE=97394457; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G., B.A.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty S.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathley L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
```

```
Venter J.C.:
"the complete genome sequence of the gastric pathogen Helicobacter
pylori.";
RL Nature 388:539-547(1997).
CC -1- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
CC EXHIBITS 3' TO 5' AND 5' TO 3' EXONUCLEASE ACTIVITY (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + (DNA)(N)
CC -1- SUBUNIT: SINGLE-CHAIN MONOMER WITH MULTIPLE FUNCTIONS.
CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-A FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000647; AAD08510.1; ALT_INIT.
CC HSP: P19821; 1TAU.
CC TIGR: HP1470.
CC InterPro: IPR002562; 3_5_exonuclease.
CC InterPro: IPR002421; 5_3_exonuclease.
CC InterPro: IPR001098; DNA_pol.
CC InterPro: IPR002298; DNA_pol.
CC InterPro: IPR000513; Exo_N.1.
CC InterPro: IPR003584; HHH.2.
CC Pfam: PF00476; DNA_pol_A.1.
CC Pfam: PF01367; 5_3_exonuclease; 1.
CC PRINTS: PR00868; DNAPOLI.
CC SMART: SM00474; 3SEXOC; 1.
CC SMART: SM00475; 5SEXOC; 1.
CC SMART: SM00279; HHH2; 1.
CC SMART: SM00482; POLAC; 1.
CC TIGRfams: TIGR00593; pole; 1.
CC PROSITE: PS00447; DNA_POLYMERASE_A; 1.
CC KW transferase; DNA-directed DNA polymerase; DNA replication; DNA repair;
CC Hydrolyase; Exonuclease; DNA-binding; Complete proteome.
CC SEQUENCE 891 AA; 101975 MW; 3D6E7D8FF613EB88 CRC64;

Query Match 1.9%; Score 111.5; DB 1; Length 891;
Best Local Similarity 18.5%; Pred. No. 11;
Matches 138; Conservative 114; Mismatches 231; Indels 261; Gaps 38;

QY 402 EKLY-STDLLPKASTGSGVRTNIYERI-----LSTLRKGFVIGDKKFEFL--AFSS-- 449
Db 213 EKIYENLDLAKNLLSPKMYRALTHDKASAFLSKELATLGRGCI---KEFDLSCAFPSEN 269
QY 450 -----SOLRD-----NSVWMEASRPGLTANDIRAMWGDFSOIKNV 484
Db 270 PLIKIKDELKEYGFTISLRDLENSPTPLIDNAPLDNTPALDNTPKKSCM-----IV 322
QY 485 AKYAARLGQSGSGSRRTLSVLRIHEIYIPDVKGVTSGYVFSGDKIGSGDF----AHRVAS 541
Db 323 LESAAPLASFLEKLEKTARVAFARLVLDKKEKVALALAFIYED----QGYPLEELALS 377
QY 542 KGLQYTPSAFQIRYGGYGVGVGVDPSDKLSRMSKSYESDNTKLDVLGSKYQPCY 601
Db 378 PFSLEFLQNAF-FKMLQHAQIIG--HDLKPLLSFLKAKYQVPLENIRIQ----- 423
QY 602 LNRQLITLLST---LGVKDEVLEQKQKQKAVDQDAILHDSLK---AQEALEMLSPGEN- 653
Db 424 -DTQILAFLNPKVGF-DEVLEKYLE-----ELPHEKIKDFKTKAEKLELLSVELNA 476
QY 654 -----TNILKAMLCN-----GYKPDAPFPLSMMLQTFR---- 681
Db 477 LKRCEYFEKGGLEENLLSLARBIETPFMKVLMGMFQGFKIDA-PYFKLEQEFKNEH 535
QY 682 -----ASKLLDLRHSRIFIPNG-----RTMMGCLDE---SRTLE 713
```

Db 536 VLQRILLEIGVDFNLNSPKOLSEVLYDKLGLPKNKSHSTDEKSLKILDKHPSIALILE 595

QY 714 YGOV-----FVQFTGAGHGFSDDLHPFNN--SRSTNSNFI 747

Db 596 YRELKLFNTYTTLLRLKDKKDDKHHTFIQ-TGTATGRLLSSHPNLQNPVRSFKGLLI 654

QY 748 LKGNVWAKNCPCLHPGDIRVLKAVNRALHVMVDCVVPQKGRKRPHPNECSGSDLDGDIY 807

Db 655 RKGFIASKEVCLLQVD---YSQTELRLLAH-----FSQ----- 685

QY 808 FVCWDQDMIPPRQVOPMEYPPAPSIQDHDVTEIEVEYFTNYIVNDSLGIIANAHVYFA 867

Db 686 ---DKDLM-----EAFKGRDHILETSKALFGEVLAKEKRSIAKSNP--- 725

QY 868 DREPDMAWSDPCKKLAELEFSTAVDFPKTGVPAEIPSLRKEYPDFDKDKTSYISERV 927

Db 726 ---GLVYGMGSKSLSTLNLSNEAKSYEAYF-----KRPSTIKD-----YLN--- 766

QY 928 IGLKFRVKDKAPQASSIATFTRDVARSYDADMEVDGFEDYIDEAFDYKTEYDNKLGNL 987

Db 767 -----RMKEEILKTSK--AFTLLGRYRVFD---FTGANDYVKG--NYLRGVNAI--- 809

QY 988 MDYVGIKTEAELSGGIMKASKTF 1011

Db 810 --FOG--SASDLLKGLMKVRSERF 829

RESULT 14

LM44\_HUMAN STANDARD; PRT; 1816 AA.

AC Q16363; Q15335; Q14735; Q9UE18; Q9JUN9;

DT 01-NOV-1997 (Rel. 35, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Laminin alpha-4 chain precursor.

GN LAMA4.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OC NCBI\_TaxID=9606;

RN [1]

RP TISSUE=Fetal lung;

RC MEDLINE=95300971; PubMed=7781776;

RA Ilvannainen A., Sainio K., Sariola H., Tryggvason K.;

RT "Primary structure and expression of a novel human laminin alpha 4 chain.";

RL FEBS Lett. 365:183-188(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Heart;

RA MEDLINE=97434279; PubMed=9310354;

RA Richards A.J., Luccarini C., Pope F.M.;

RT "The structural organisation of LAMA4, the gene encoding laminin alpha4.";

RL Eur. J. Biochem. 248:15-23(1997).

RN [3]

RP SEQUENCE OF 236-1816 FROM N.A.

RC TISSUE=Heart;

RA MEDLINE=95048381; PubMed=7959779;

RA Richards A.J., Al-Imara L., Carter N.P., Lloyd J.C., Leversha M.A., Pope F.M.;

RT "Localization of the gene (LAMA4) to chromosome 6q21 and isolation of a partial cDNA encoding a variant laminin A chain.";

RL Genomics 22:237-239(1994).

RN [4]

RP SEQUENCE OF 66-1816 FROM N.A.

RA Tubby B.;

RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin is thought to mediate the attachment, migration, and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components.

CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three different polypeptide chains (alpha, beta, gamma), which are bound to each other by disulfide bonds into a cross-shaped molecule comprising one long and three short arms with globules at each end.

CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT MEMBRANES (MAJOR COMPONENT).

CC -!- TISSUE SPECIFICITY: IN ADULT, STRONG EXPRESSION IN HEART, LUNG, OVARY SMALL AND LARGE INTESTINES, PLACENTA, LIVER; WEAK OR NO EXPRESSION IN SKELETAL MUSCLE, KIDNEY, PANCREAS, TESTIS, PROSTATE, BRAIN. HIGH EXPRESSION IN FETAL LUNG AND KIDNEY. EXPRESSION IN FETAL AND NEWBORN TISSUES IS OBSERVED IN CERTAIN MESENCHYMAL CELLS IN TISSUES SUCH AS SMOOTH MUSCLE AND DERMIS.

CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.

CC -!- DOMAIN: DOMAIN G IS GLOBULAR.

CC -!- SIMILARITY: CONTAINS 3-5 LAMININ EGF-LIKE DOMAINS.

CC -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.

CC -!- CAUTION: GENE LAMA4 WAS FORMERLY CALLED LAMA3.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; S78569; AAB34635.1; -

DR EMBL; X91171; CAB62596.1; -

DR EMBL; Y14240; CAAY4636.1; -

DR EMBL; X76939; CAA54258.1; -

DR EMBL; Z99289; CAB16553.1; -

DR HSSP; P02468; 1KLO.

DR Genew; HGNC:6484; LAMA4.

DR MIM; 600133; -

DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR002049; Laminin\_EGF.

DR InterPro; IPR001791; Laminin\_G.

DR Pfam; PF00053; laminin\_EGF; 6.

DR Pfam; PF00054; laminin\_G; 6.

DR SMART; SM00180; EGF\_Lam; 3.

DR SMART; SM00282; LamG; 5.

DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.

DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 3.

DR PROSITE; PS50025; LAM\_G\_DOMAIN; 5.

KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil; Laminin EGF-like domain; Cell adhesion; Repeat; Signal.

FT SIGNAL 1 24

FT CHAIN 25 1816 LAMININ ALPHA-4 CHAIN.

FT DOMAIN 82 131 LAMININ EGF-LIKE 1.

FT DOMAIN 132 186 LAMININ EGF-LIKE 2.

FT DOMAIN 187 240 LAMININ EGF-LIKE 3.

FT DOMAIN 241 255 LAMININ EGF-LIKE 4 (INCOMPLETE).

FT DOMAIN 256 825 DOMAIN II AND I.

FT DOMAIN 826 1028 LAMININ G-LIKE 1.

FT DOMAIN 1040 1220 LAMININ G-LIKE 2.

FT DOMAIN 1227 1395 LAMININ G-LIKE 3.

FT DOMAIN 1462 1633 LAMININ G-LIKE 4.

FT DOMAIN 1640 1813 LAMININ G-LIKE 5.

FT DOMAIN 313 396 COILED COIL (POTENTIAL).

FT DOMAIN 466 521 COILED COIL (POTENTIAL).

FT DOMAIN 574 607 COILED COIL (POTENTIAL).

FT DOMAIN 655 717 COILED COIL (POTENTIAL).

FT DOMAIN 770 799 COILED COIL (POTENTIAL).

FT SITE 717 719 CELL ATTACHMENT SITE (POTENTIAL).

FT DISULFID 82 91 BY SIMILARITY.

FT DISULFID 84 98 BY SIMILARITY.

FT DISULFID 101 110 BY SIMILARITY.

FT DISULFID 113 129 BY SIMILARITY.

FT DISULFID 132 146 BY SIMILARITY.

FT DISULFID 134 155 BY SIMILARITY.

FT DISULFID 157 166 BY SIMILARITY.





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:42:08 ; Search time 40 Seconds  
(without alignments)  
5738.412 Million cell updates/sec

Title: US-09-782-874-2

Perfect score: 5864

Sequence: 1 MGKTIQVGFPPYLLSAEVVK.....RPVLNLSLRAQLSHRLVLK 1114

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP\_TREMBL\_21.\*  
2: SP\_Archaea.\*  
3: SP\_Bacteria.\*  
4: SP\_Fungi.\*  
5: SP\_Human.\*  
6: SP\_Invertebrate.\*  
7: SP\_Mammal.\*  
8: SP\_MHC.\*  
9: SP\_Organelle.\*  
10: SP\_Phage.\*  
11: SP\_Plant.\*  
12: SP\_Rodent.\*  
13: SP\_Virus.\*  
14: SP\_Vertebrate.\*  
15: SP\_Unclassified.\*  
16: SP\_Virus.\*  
17: SP\_Bacteriophage.\*  
17: SP\_Archaea.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5864	100.0	1114	10	Q9Z58 lycopersico
2	5070	86.5	1116	10	Q9ZRY7 nicotiana t
3	3598.5	61.4	775	10	Q9ZRX3 petunia hyb
4	3532	60.2	1107	10	Q9LQV2 arabidopsis
5	1995.5	34.0	1133	10	Q82504 arabidopsis
6	1984	33.8	513	10	Q9ZRX5 arabidopsis
7	1621	27.6	1196	10	Q9SG02 arabidopsis
8	1610	27.5	1196	10	Q9LKP0 arabidopsis
9	1088.5	18.2	1638	3	Q9C162 neurospora
10	1031	17.6	1484	3	Q8TG4 diaportha p
11	995	17.0	1483	3	Q96UL6 diaportha a
12	987	16.8	1122	3	Q8XIC2 phomopsis s
13	948	16.2	1215	3	O14227 schizosacch
14	932	15.9	1780	5	O19285 caenorhabdi
15	898	15.3	1579	5	Q9B56 caenorhabdi
16	896.5	15.3	1601	5	Q9NDH1 caenorhabdi

17 860.5 14.7 1632 5 Q93593  
18 720 12.3 1964 5 Q9ZG6  
19 703 1883 5 Q9ZG6  
20 515.5 8.8 1069 5 Q9XYPI  
21 456 7.8 107 10 Q9ZRX4  
22 441 7.5 953 3 Q01869  
23 385.5 6.6 1377 5 Q95WU3  
24 362 6.2 966 10 Q82190  
25 355.5 6.1 1148 10 Q94IV7  
26 350.5 6.0 929 10 Q82188  
27 345.5 5.9 1222 10 Q94IV6  
28 335.5 5.7 905 10 Q82189  
29 297 5.1 1402 3 Q9Y7G6  
30 145.5 2.5 634 4 Q96NH6  
31 140 2.4 2029 5 Q9VD07  
32 138.5 2.4 1416 4 Q9BZF9  
33 138.5 2.4 1449 6 Q9BGF7  
34 134.5 2.3 549 3 Q9USN8  
35 134 2.3 1308 12 Q91DN2  
36 134 2.3 1816 12 Q91DN3  
37 133.5 2.3 2364 2 Q46342  
38 133 2.3 2810 5 Q20456  
39 132 2.3 3072 12 Q92645  
40 130.5 2.2 1091 16 Q9KAQ7  
41 130 2.2 781 16 Q8XL45  
42 129.5 2.2 849 10 Q9SA72  
43 129 2.2 804 10 Q8W3J7  
44 128.5 2.2 839 9 Q64076  
45 128.5 2.2 839 16 Q31945

#### ALIGNMENTS

#### RESULT 1

Q9Z58 PRELIMINARY; PRT; 1114 AA.  
ID Q9Z58  
AC Q9Z58  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE RNA-directed RNA polymerase (EC 2.7.7.48).  
GN RDRP.  
OS Lycopersicon esculentum (Tomato).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=4081;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RUTGERS;  
RX MEDLINE=99055198; PubMed=9836747;  
RA Schiebel W., Fellissier T., Riedel L., Thalmair S., Schiebel R.,  
RT Kempe D., Lottspeich F., Sanger H.L., Wassenegeger M.;  
RT "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from  
RT tomato.";  
RL Plant Cell 10:2087-2102(1998).  
DR EMBL; Y10403; CAA71421.1; -;  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR InterPro; IPR003955; RRM\_2.  
DR Pfam; PF00076; rrm; 1.  
DR SMART; SM00362; RRM\_3; 1.  
KW Nucleotidyltransferase; RNA-directed RNA polymerase; Transferase.  
SQ SEQUENCE 1114 AA; 126807 MW; B6D30ACD41DB37FF CRC64;

Query Match 100.0%; Score 5864; DB 10; Length 1114;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MGKTIQVGFPPYLLSAEVVKSFLEKTYGTGTCVALEVKQSGSRAFAKVQFADNISADK 60  
Db 1 MGKTIQVGFPPYLLSAEVVKSFLEKTYGTGTCVALEVKQSGSRAFAKVQFADNISADK 60

QY	61	IITLANNRLFGSSYLKAWEMKTDIVQLRAYVDQMDGIIITNFGCOISDDKFAVLGSTEVS	120
Db	61	IITLANNRLFGSSYLKAWEMKTDIVQLRAYVDQMDGIIITNFGCOISDDKFAVLGSTEVS	120
QY	121	IQFGIGLKKFFFLSSGSADYKQLQSYENITWQVYLHRYPGQNAQFLLIQLFGAPRIYKRL	180
Db	121	IQFGIGLKKFFFLSSGSADYKQLQSYENITWQVYLHRYPGQNAQFLLIQLFGAPRIYKRL	180
QY	181	ENSCYSFFKETPDQWVRTDFFPSWITGLSSSLCQLQPRGRVRLPNFEESFFHYAERENNI	240
Db	181	ENSCYSFFKETPDQWVRTDFFPSWITGLSSSLCQLQPRGRVRLPNFEESFFHYAERENNI	240
QY	241	TLOGFTFFFSQKALPNVQPPGEGISPYKILFKISSLVHQCICPGPALNVYFFRLVDP	300
Db	241	TLOGFTFFFSQKALPNVQPPGEGISPYKILFKISSLVHQCICPGPALNVYFFRLVDP	300
QY	301	RRNVACIEHALEKLYIKKCCDVPVRLTEQYDGYLKGROPKSPSITLDDGLVYVRRV	360
Db	301	RRNVACIEHALEKLYIKKCCDVPVRLTEQYDGYLKGROPKSPSITLDDGLVYVRRV	360
QY	361	LVPCKYVFCGPEVNVNRVLRNYSEDIDNFLRVSFVDEWEKLYSTDLLPKASTGSGVR	420
Db	361	LVPCKYVFCGPEVNVNRVLRNYSEDIDNFLRVSFVDEWEKLYSTDLLPKASTGSGVR	420
QY	421	TNIYERILSTRKGFVIGDKKFFELAFSSQLRDNVSMFASRPGLTANDIRAWMGDFSQ	480
Db	421	TNIYERILSTRKGFVIGDKKFFELAFSSQLRDNVSMFASRPGLTANDIRAWMGDFSQ	480
QY	481	IKNVAKYAARLGQSGSSRETLSVLRIHEIEVIPDVKHGTSYVFSOGIGKISGDFAHRYA	540
Db	481	IKNVAKYAARLGQSGSSRETLSVLRIHEIEVIPDVKHGTSYVFSOGIGKISGDFAHRYA	540
QY	541	SKCGLQTPSAFQIRYGGYKGVVGVDPDSSMKLSLRKSMKSYESDNIKLDVLGWSKYQPC	600
Db	541	SKCGLQTPSAFQIRYGGYKGVVGVDPDSSMKLSLRKSMKSYESDNIKLDVLGWSKYQPC	600
QY	601	YLNQLTLITLSTLGKDEVLEQKQKAVDQDAILHDSLKAQALELMSPGENTNLIKAM	660
Db	601	YLNQLTLITLSTLGKDEVLEQKQKAVDQDAILHDSLKAQALELMSPGENTNLIKAM	660
QY	661	LNCGYKPDABPFLSMIQTFRASKLLDLRTSRIFIPNGRTMGCIDESTLEYGVQVFO	720
Db	661	LNCGYKPDABPFLSMIQTFRASKLLDLRTSRIFIPNGRTMGCIDESTLEYGVQVFO	720
QY	721	FTGAGHGEFSDLLHFPFNNSRSTNSNFIKGNVYVAKNCPCLHPGDIIRYLKAVNRYALHHMV	780
Db	721	FTGAGHGEFSDLLHFPFNNSRSTNSNFIKGNVYVAKNCPCLHPGDIIRYLKAVNRYALHHMV	780
QY	781	DCVVEPKGKRPHNECGSDLDGDIYFVCWDQDMIPPROVQPMVEYPPAPSIOLDHDVTI	840
Db	781	DCVVEPKGKRPHNECGSDLDGDIYFVCWDQDMIPPROVQPMVEYPPAPSIOLDHDVTI	840
QY	841	BEVEEYETNIVNDSLGIIANAHVVFADRPDMAMSDPCCKLAEPLSIAVDFFPKTGVPAE	900
Db	841	BEVEEYETNIVNDSLGIIANAHVVFADRPDMAMSDPCCKLAEPLSIAVDFFPKTGVPAE	900
QY	901	IPSOLRKEYPDFMDKPDKTSYISERVIGKLFKVKDKDKAPQASSIAFTFEDVARRSYDAD	960
Db	901	IPSOLRKEYPDFMDKPDKTSYISERVIGKLFKVKDKDKAPQASSIAFTFEDVARRSYDAD	960
QY	961	MEVDGFEDYIDEAFDYKTEYDNKLNLMYDYGITKTEAELISGGIMKASKTFDRRKDAEAI	1020
Db	961	MEVDGFEDYIDEAFDYKTEYDNKLNLMYDYGITKTEAELISGGIMKASKTFDRRKDAEAI	1020
QY	1021	SVAVRALRKARAWFKRRNDIDMLPKASAWYHVHTYWGTCYNOGLKRAHFISPPWCY	1080
Db	1021	SVAVRALRKARAWFKRRNDIDMLPKASAWYHVHTYWGTCYNOGLKRAHFISPPWCY	1080
QY	1081	YDQLIQIKKDKARNRPNVNLSSSLRAQLSHRLVLK 1114	
Db	1081	YDQLIQIKKDKARNRPNVNLSSSLRAQLSHRLVLK 1114	

QY	61	IITLANNRLFGSSYLKAWEMKTDIVQLRAYVDQMDGIIITNFGCOISDDKFAVLGSTEVS	120
Db	61	IITLANNRLFGSSYLKAWEMKTDIVQLRAYVDQMDGIIITNFGCOISDDKFAVLGSTEVS	120
QY	121	IQFGIGLKKFFFLSSGSADYKQLQSYENITWQVYLHRYPGQNAQFLLIQLFGAPRIYKRL	180
Db	121	IQFGIGLKKFFFLSSGSADYKQLQSYENITWQVYLHRYPGQNAQFLLIQLFGAPRIYKRL	180
QY	181	ENSCYSFFKETPDQWVRTDFFPSWITGLSSSLCQLQPRGRVRLPNFEESFFHYAERENNI	240
Db	181	ENSCYSFFKETPDQWVRTDFFPSWITGLSSSLCQLQPRGRVRLPNFEESFFHYAERENNI	240
QY	241	TLOGFTFFFSQKALPNVQPPGEGISPYKILFKISSLVHQCICPGPALNVYFFRLVDP	300
Db	241	TLOGFTFFFSQKALPNVQPPGEGISPYKILFKISSLVHQCICPGPALNVYFFRLVDP	300
QY	301	RRNVACIEHALEKLYIKKCCDVPVRLTEQYDGYLKGROPKSPSITLDDGLVYVRRV	360
Db	301	RRNVACIEHALEKLYIKKCCDVPVRLTEQYDGYLKGROPKSPSITLDDGLVYVRRV	360
QY	361	LVPCKYVFCGPEVNVNRVLRNYSEDIDNFLRVSFVDEWEKLYSTDLLPKASTGSGVR	420
Db	361	LVPCKYVFCGPEVNVNRVLRNYSEDIDNFLRVSFVDEWEKLYSTDLLPKASTGSGVR	420
QY	421	TNIYERILSTRKGFVIGDKKFFELAFSSQLRDNVSMFASRPGLTANDIRAWMGDFSQ	480
Db	421	TNIYERILSTRKGFVIGDKKFFELAFSSQLRDNVSMFASRPGLTANDIRAWMGDFSQ	480
QY	481	IKNVAKYAARLGQSGSSRETLSVLRIHEIEVIPDVKHGTSYVFSOGIGKISGDFAHRYA	540
Db	481	IKNVAKYAARLGQSGSSRETLSVLRIHEIEVIPDVKHGTSYVFSOGIGKISGDFAHRYA	540
QY	541	SKCGLQTPSAFQIRYGGYKGVVGVDPDSSMKLSLRKSMKSYESDNIKLDVLGWSKYQPC	600
Db	541	SKCGLQTPSAFQIRYGGYKGVVGVDPDSSMKLSLRKSMKSYESDNIKLDVLGWSKYQPC	600
QY	601	YLNQLTLITLSTLGKDEVLEQKQKAVDQDAILHDSLKAQALELMSPGENTNLIKAM	660
Db	601	YLNQLTLITLSTLGKDEVLEQKQKAVDQDAILHDSLKAQALELMSPGENTNLIKAM	660
QY	661	LNCGYKPDABPFLSMIQTFRASKLLDLRTSRIFIPNGRTMGCIDESTLEYGVQVFO	720
Db	661	LNCGYKPDABPFLSMIQTFRASKLLDLRTSRIFIPNGRTMGCIDESTLEYGVQVFO	720
QY	721	FTGAGHGEFSDLLHFPFNNSRSTNSNFIKGNVYVAKNCPCLHPGDIIRYLKAVNRYALHHMV	780
Db	721	FTGAGHGEFSDLLHFPFNNSRSTNSNFIKGNVYVAKNCPCLHPGDIIRYLKAVNRYALHHMV	780
QY	781	DCVVEPKGKRPHNECGSDLDGDIYFVCWDQDMIPPROVQPMVEYPPAPSIOLDHDVTI	840
Db	781	DCVVEPKGKRPHNECGSDLDGDIYFVCWDQDMIPPROVQPMVEYPPAPSIOLDHDVTI	840
QY	841	BEVEEYETNIVNDSLGIIANAHVVFADRPDMAMSDPCCKLAEPLSIAVDFFPKTGVPAE	900
Db	841	BEVEEYETNIVNDSLGIIANAHVVFADRPDMAMSDPCCKLAEPLSIAVDFFPKTGVPAE	900
QY	901	IPSOLRKEYPDFMDKPDKTSYISERVIGKLFKVKDKDKAPQASSIAFTFEDVARRSYDAD	960
Db	901	IPSOLRKEYPDFMDKPDKTSYISERVIGKLFKVKDKDKAPQASSIAFTFEDVARRSYDAD	960
QY	961	MEVDGFEDYIDEAFDYKTEYDNKLNLMYDYGITKTEAELISGGIMKASKTFDRRKDAEAI	1020
Db	961	MEVDGFEDYIDEAFDYKTEYDNKLNLMYDYGITKTEAELISGGIMKASKTFDRRKDAEAI	1020
QY	1021	SVAVRALRKARAWFKRRNDIDMLPKASAWYHVHTYWGTCYNOGLKRAHFISPPWCY	1080
Db	1021	SVAVRALRKARAWFKRRNDIDMLPKASAWYHVHTYWGTCYNOGLKRAHFISPPWCY	1080
QY	1081	YDQLIQIKKDKARNRPNVNLSSSLRAQLSHRLVLK 1114	
Db	1081	YDQLIQIKKDKARNRPNVNLSSSLRAQLSHRLVLK 1114	

RESULT 2

Q9ZRY7

PRELIMINARY; PRT: 1116 AA.

AC Q9ZRY7;

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE RNA-directed RNA polymerase.

GN RDRP.

OS Nicotiana tabacum (Common tobacco).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.

OX NCBI\_TaxID=4097;

RN [1]

SEQUENCE FROM N.A.

RC STRAIN=CV. PETIT HAVANA SRI; TISSUE=LEAF;

RA Schiebel W., Pellissier T., Riedel L., Thalmeir S., Schiebel R.,

RA Kempe D., Lottspeich F., Saenger H.L., Wassenecker M.;

RT "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from

RT Tomato";

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ011576; CAA09697.1; -.

KW RNA-directed RNA polymerase.

SQ SEQUENCE 1116 AA; 127553 MW; 8D9889DA826D31DB CRC64;

Query Match

Best Local Similarity 86.5%; Score 5070; DB 10; Length 1116;

Mismatches 959; Conservative 67; Mismatches 87; Indels 2; Gaps 1;

QY 1 MCKTIQVFGFPYLLSAEVVKSLEYKTYGTVCALELVKQSGKGRAFKQVADMISADK 60

Db 1 MCKTIQVFGFPYLLSAEVVKSLEYKTYGTVCALELVKQSGKGRAFKQVADMISADK 60

QY 61 IITLANNRLFGSSYLKAWEMKTDIVQLRAYVDQMDGIIITNFGCOISDDKFAVLGSTEVS 120

Db 61 IITLANNRLFGSSYLKAWEMKTDIVQLRAYVDQMDGIIITNFGCOISDDKFAVLGSTEVS 120

QY 121 IQFGIGLKKFFFLSSGSADYKQLQSYENITWQVYLHRYPGQNAQFLLIQLFGAPRIYKRL 180

Db 121 IQFGIGLKKFFFLSSGSADYKQLQSYENITWQVYLHRYPGQNAQFLLIQLFGAPRIYKRL 180

QY 181 ENSCYSFFKETPDQWVRTDFFPSWITGLSSSLCQLQPRGRVRLPNFEESFFHYAERENNI 240

Db 181 ENSCYSFFKETPDQWVRTDFFPSWITGLSSSLCQLQPRGRVRLPNFEESFFHYAERENNI 240

QY 241 TLOGFTFFFSQKALPNVQPPGEGISPYKILFKISSLVHQCICPGPALNVYFFRLVDP 300

Db 241 TLOGFTFFFSQKALPNVQPPGEGISPYKILFKISSLVHQCICPGPALNVYFFRLVDP 300

QY 301 RRRNVACIEHALEKLYIKKCCDVPVRLTEQYDGYLKGROPKSPSITLDDGLVYVRRV 360

Db 301 RRRNVACIEHALEKLYIKKCCDVPVRLTEQYDGYLKGROPKSPSITLDDGLVYVRRV 360

QY 361 LVPCKYVFCGPEVNVNRVLRNYSEDIDNFLRVSFVDEWEKLYSTDLLPKASTGSGVR 420

Db 361 LVPCKYVFCGPEVNVNRVLRNYSEDIDNFLRVSFVDEWEKLYSTDLLPKASTGSGVR 420

QY 421 TNIYERILSTRKGFVIGDKKFFELAFSSQLRDNVSMFASRPGLTANDIRAWMGDFSQ 480

Db 421 TNIYERILSTRKGFVIGDKKFFELAFSSQLRDNVSMFASRPGLTANDIRAWMGDFSQ 480

QY 481 IKNVAKYAARLGQSGSSRETLSVLRIHEIEVIPDVKHGTSYVFSOGIGKISGDFAHR 538

Db 481 IKNVAKYAARLGQSGSSRETLSVLRIHEIEVIPDVKHGTSYVFSOGIGKISGDFAHR 540

QY 539 VASKGLQYTPSAFQIRYGGYKGVVGVDPDSSMKLSLRKSMKSYESDNIKLDVLGWSKYQ 598

Db 541 VTIKGGLQYTPSSQIRYGGYKGVVGVDPDSSMKLSLRKSMKSYESDNIKLDVLGWSKYQ 600

QY 599 PCYLNROLITLSTLGKDEVLEQKQKAVDQDAILHDSLKAQALELMSPGENTNLIK 658

Db 601 PCYLNROLITLSTLGKDEVLEQKQKAVDQDAILHDSLKAQALELMSPGENTNLIK 660

```
QY 659 AMLNCGYKPDAPFLSMMLQTFRASKLLDLTRSRIFIPNGRTMMGCLDESRTLEYQGVF 718
Db 661 EMLNCGYMPDAEPFLSMMLQTFRASKLLDLTRSRIFIPNGRTMMGCLDESRTLEYQGVF 720
QY 719 VQFTGAGHGFSDDLHPFNRSRNSNFILKGNVVAKNPCLHPGDIRVLKAVNRALHH 778
Db 721 VQFSGARRQFFESHFNFSGSANCDFILKGNVVAKNPCLHPGDIRVLAVDVPALHH 780
QY 779 MVDGWFFQKRPHPNECSGSLDGDIIYFVNDQDMIPPRQVQPMPEYPPAPSIQLDHDV 838
Db 781 MVDGWFFQKRPHPNECSGSLDGDIIYFVNDQDLPPIPRQVQSMYDTPAPITQLDHDV 840
QY 839 TIEVEEYFNINVDLSGLIIANAHVVFADREPDMAMSDPCKLAELFSTAVDPKTVGP 898
Db 841 TIEVEEYFNINVDLSGLIIANAHVVFADREPDMAMSDPCKLAELFSTAVDPKTVGP 900
QY 899 AEIPSQLRPXEPDMEKPKPTYPSEKVIKLFQKVNKTQPASSIATFTRVARKSYD 958
Db 901 AEIPSQLRPXEPDMEKPKPTYPSEKVIKLFQKVNKTQPASSIATFTRVARKSYD 960
QY 959 ADMEVDGFEDYIDAFYKTEYDNKGLNMDYGIKTEAEILSGGINKASKTFDRRKDAE 1018
Db 961 SDMTVDGFEDYIDAFYKTEYDNKGLNMDYGIKTEAEILSGGINKASKTFDRRKDAE 1020
QY 1019 AISVAVRALRKEARAWFKRRNDIDMLPKASAWHYVTHYHTYGCYNQGLKRAHFISFPW 1078
Db 1021 AIGVAVCLRKEARAWFKRRNDIDMLAKASAWHYVTHYHTYGCYNQGLKRAHFISFPW 1080
QY 1079 CVYDQLIOIKDKARNRPNVLSLRAQLSHRLVL 1113
Db 1081 CVYDQLIOIKKAKRKPVPHELSSIGSOLSKLVI 1115

RESULT 3
Q92RX3 ID Q92RX3 PRELIMINARY; PRT; 775 AA.
AC Q92RX3;
DT 01-MAY-1999 (TremBLrel. 10, Created)
DT 01-MAY-1999 (TremBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE RNA-directed RNA polymerase (fragment).
GN RDRP.
OS Petunia hybrida (Petunia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
OX NCBI_TaxID=4102;
RN [1];
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Schiebel W., Pelissier T., Riedel L., Thalmair S., Schiebel R.,
RA Kempe D., Lottspeich F., Saenger H.L., Wassenegger M.;
RT "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from
RT Tomato.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011979; CAA09896.1; -
KW RNA-directed RNA polymerase.
FT NON_TER 1
FT NON_TER 775
SQ SEQUENCE 775 AA; 88778 MW; DE43FE50C6A45B93 CRC64;

Query Match 61.4%; Score 3598.5; DB 10; Length 775;
Best Local Similarity 87.0%; Pred. No. 7.7e-251;
Matches 675; Conservative 47; Mismatches 53; Indels 1; Gaps 1;

QY 307 CIEHALEKLYIKCCYDPPRWLTQDGYLKGKQPPKSPSITLDGLVVRVRLVTPCK 366
Db 1 CIEHALEKLYIKCCYDPPRWLTQDGYLKGKQPPKSPSITLDGLVVRVRLVTPCK 60
QY 367 VYFGPEVNVNRYLRNYSIDINFLRVSVDEWEKLYSTDLLPKASTGSGVRTNIYER 426
Db 61 VYFLGPEVNVNRYLRNYSIDINFLRVSVDEWEKLYSTDLLPKASTGSGVRTNIYER 119
```

```
QY 427 ILSTLRKGEVIGDKFEFLAFSSQLRDRNSVVMFASRPGLTANDIRAMWGDFSQIKNVAK 486
Db 120 ILSTLQNLVIGAKKFEFLAFSSQLRDRNSVVMFASRPGLTAYDIRTWGDFQOIRNVAK 179
QY 487 YAARLQSGSSRRETLVLRHTEVTPDVKHGTSYVFSDDGIGKISGDFAHFVASKCGLQ 546
Db 180 YAARLQSGSSRRETLVSRHTEVTPDVKHGTSYVFSDDGIGKISGDFAHFVASKCGLQ 239
QY 547 YTPSAFQIRYGGYKGVVGVDPDSSMKLSLRKSMKSKYEDSNILKDLVLGWSKYQPCYLNRL 606
Db 240 HTPSAFQIRYGGYKGVVGVDPDSSMKLSLRKSMKSKYEDSNILKDLVLGWSKYQPCYLNRL 299
QY 607 IYLLSTLGVKDEVLQKQKAEVNDQDAIILHDSLKQAELSLMSPGENTILKAMLCNGYK 666
Db 300 VTLMSLTVGVKDEVLQKQKAEVNDQDAIILHDSLKQAELSLMSPGENTILKAMLCNGYK 359
QY 667 PDAEPFLSMMLQTFRASKLLDLTRSRIFIPNGRTMMGCLDESRTLEYQGVFQFTGAGH 726
Db 360 PDAEPFLSMMLQTFRASKLLDLTRSRIFIPNGRTMMGCLDESRTLEYQGVFQFTGAGH 419
QY 727 GFESDDLHPFNRSRNSNFILKGNVVAKNPCLHPGDIRVLKAVNRALHHMDCVVP 786
Db 420 RQFYEESEHFNFSGSANYNFILKGNVVAKNPCLHPGDIRVLKAVDVPALHHMDCVVP 479
QY 787 QXGKRPHNECSGSLDGDIIYFVNDQDMIPPRQVQPMPEYPPAPSIQLDHDVTEIEVEY 846
Db 480 QXGKRPHNECSGSLDGDIIYFVNDQDLPPIPRQVQPMPEYPPAPSIQLDHDVTEIEVEY 539
QY 847 FTYNIYNDLSGLIIANAHVVFADREPDMAMSDPCKLAELFSTAVDPKTVGPAPISQRL 906
Db 540 FTYNIYNDLSGLIIANAHVVFADREPDMAMSDPCKLAELFSTAVDPKTVGPAPISQRL 599
QY 907 PREYDPFMDKPKTSYISERVIGKLFKRYKDKAPQASSIATFTRVARSYDADMEYDGF 966
Db 600 PREYDPFMDKPKTSYISERVIGKLFKRYKDKAPQASSIATFTRVARSYDADMEYDGF 659
QY 967 EDYIDEAFDYKTEYDNKGLNMDYGIKTEAEILSGGINKASKTFDRRKDAEISAVRA 1026
Db 660 EDYIDEAFDYKTEYDNKGLNMDYGIKTEAEILSGGINKASKTFDRRKDAEISAVRA 719
QY 1027 LRKEARAWFKRRNDIDMLPKASAWHYVTHYHTYGCYNQGLKRAHFISFPWCYVD 1082
Db 720 LRKEARAWFKRRNDIDMLAKASAWHYVTHYHTYGCYNQGLKRAHFISFPWCYVD 775

RESULT 4
Q9LOV2 ID Q9LOV2 PRELIMINARY; PRT; 1107 AA.
AC Q9LOV2;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE F1086.19.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1];
RP SEQUENCE FROM N.A.
RA Chao Q., Shinn P., Dunn P., Buehler E., Kahn S., Kim C., Walker M.,
RA Williams S., Altairi H., Araujo R., Conn L., Conway A.B., Gonzalez A.,
RA Hansen N.F., Huizar L., Kremenetskaia I., Lenz C., Li J., Liu S.,
RA Luros S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V., Yu G.,
RA Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F10B6 from chromosome
RT I.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2];
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
```



```
[3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN
[4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
  Khan S., Kim C., Altati H., Bei B., Chin C., Chou J., Choi E.,
  Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
  Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharasy N.,
  Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
  Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
  Theologis A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006917; AAF79241.1; -
SQ
  Query Match 60.2%; Score 3532; DB 10; Length 1107;
  Best Local Similarity 62.4%; Pred. No. 8.6e-246;
  Matches 692; Conservative 160; Mismatches 225; Indels 32; Gaps 12;
  QY 1 MGKTIQVGFPPYLLSAEYVKSFLKTYGTGTCALVEYKQ-SKGSRAFAKVPQADNISAD 59
  Db 1 MGKTIQVGFPPYVSAEYVKSFLKTYGTGTCALVEYKQ-SKGSRAFAKVPQADNISAD 60
  QY 60 KIITLANNRLVFGSSYLKAMEMKTDIV-OLRAYVDQMDGITLNFQGLSDDKFAVLGST- 117
  Db 61 LIITAAERLYIGRSYLKAFEEVQDIPKPRASLHTISGLKMFQGCVOVSKFKLTLSAQ 120
  QY 118 EVSIQFGLKFFFFSSSADYKLOLSYENIWOVYLHRPYQNAOFLLIQFQAPRIY 177
  Db 121 DVCVSFGIGMKLHFSWYKQYRLESLYENIWOIDLHSPQGRSSFLYIVQIGAPKIF 180
  QY 178 KLEN-----SCYSFPEKPEDDQWVTTDF-PPSWIGLSSSLCQPRGRVRLPNEESF 230
  Db 181 EKEDQINLFLGIMDFYSDGSDEQWITDTFTSSSCIGQSTAFCLFVPLVHLNVPDFRNF 240
  QY 231 PHYAE-RENNITLTQFTFFYSQKSAVNPVQPEGISIPYKILFKLSSVQHGCIQPGA 289
  Db 241 ANYAEHRASSFLBSG-SYSSNANTLVVPPPPGSLPFEILKLTILVQNAQLSGPA 299
  QY 290 LNVFFRLVDPNRNAVACIEHALKLYIKCCYDPVWLTQYDGLKGRQPKPSIT 349
  Db 300 LDLDYFLNKKYDRALIDHCKLEKFLGECCEYEPAAHLRDEYKWKISGKLPLSPSTIS 359
  QY 350 LDDGLVYVRRVLTVPCKYFCGPPVNVSNVRLNYSIDIDNFRVSPVDEWEKLYSTD 409
  Db 360 LDDGLVYMRVQVTPARVYFSGPPVNVSNVRLNYSIDIDNFRVSPVDEWEKLYSTD 419
  QY 410 LPKASTGSGVRTNIYERILSTRKGFVIGDKKFEFLAFSSQLRDSNVWFMFASRGLTAN 469
  Db 420 SPRSGTOR--RTKLDYRIYSVLRDGVIGDKKFEFLAFSSQLRDSNVWFMFASRGLTAN 477
  QY 470 DIRAMGDFSOIKNVAKYARLGOSFSSRETLSVLRHEIEVTPDKV--HGTYSYFSDG 527
  Db 478 HIRAMGDFHIRNVAKYARLGOSFSSRETLSVLRHEIEVTPDKV--HGTYSYFSDG 537
  QY 528 ICKISGDFAHVASKGL-QYTPSAFQIRYGGYGVYVDPDSSMKLSLRKSMKSYESDN 586
  Db 538 ICKISGDFAHVASKGL-QYTPSAFQIRYGGYGVYVDPDSSMKLSLRKSMKSYESDN 597
  QY 587 IKDLVGLWSKYQPCYLNRLQTLITLSTLVGKDEVLQKQKAVDOLDAILHDSLKAQPALE 646
  Db 598 TKDLVLAWSKYQPCYLNRLQTLITLSTLVGKDEVLQKQKAVDOLDAILHDSLKAQPALE 657
  QY 647 LMSPGENTINILKAMNGCYKPDAPFFLSMMLQTPFRASKLLDLTRSRIFIPNGRTMMGCL 706
  Db 658 LMAFGENTINILKALILCYKPDAPFFLSMMLQTPFRASKLLDLTRSRIFIPNGRTMMGCL 717
  QY 707 DESTLEYGVVQVQFTGAGHGEFSDDLHPFNNSRSTNSNFIKGNVVAKNPCLHPGDIR 766
  Db 718 DETRTLEYGVVQVQ-----YSDPMRP-----GRRFIITGPPVVAKNPCLHPGDVR 762
```

```
767 VLKAVNVRALHHMVCVFPQKGRPHNECSGSLDGDYFVWCWDQDMIPRQVQPMY 826
  Db 763 VLQAVNVRALHHMVCVFPQKGRPHNECSGSLDGDYFVWCWDQELVPPRTSEPMY 822
  QY 827 PRAPSQLDHVDVTIEVEEYFTNYIVNDSLGIIAHNVFADREPDMDSDCKKLAE 886
  Db 823 TPEPTQILDHVDVTIEVEEYFTNYIVNDSLGIIAHNVFADREPDMDSDCKKLAE 882
  QY 887 STAVDPFKTGVPAEIPSQLRPEYDFMDPKDTSVISERVIGKLFKRVKDKAPQAS 946
  Db 883 STAVDPFKTGVPAEIPSQLRPEYDFMDPKDTSVISERVIGKLFKRVKDKAPQAS 942
  QY 947 TTRDVARSYDADMEVDGFEDYIDEAFYKTEYNKLNMDYIGIKTEAEILSGGIM 1006
  Db 943 SETLDVASKSYDKMEVDGFEDYIDEAFYKTEYNKLNMDYIGIKTEAEILSGGIM 1002
  QY 1007 ASKTFORRRKDAEISAVVRAALRKEARWFRNRDDDDMLPKASAWHYHTYHTY 1066
  Db 1003 MSKSFTRKDAEISAVVRAALRKEARWFRNRDDDDMLPKASAWHYHTYHTY 1062
  QY 1067 GLKRAHFISFPWCYDQLIQIKKDKARNR 1095
  Db 1063 GLNRDHFSLFAMCVYDKLVRIKKTNLGR 1091
```

```
RESULT 5
O82504 PRELIMINARY; PRT; 1133 AA.
AC O82504;
DT 01-NOV-1998 (TREMREL. 08, Created)
DT 01-NOV-1998 (TREMREL. 08, Last sequence update)
DT 01-NOV-2002 (TREMREL. 20, Last annotation update)
DE F2P3.11 protein (putative RNA-directed RNA polymerase).
GN F2P3.11 OR T22B4.110 OR AT4G11130.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA WASHU;
RT "The A. thaliana Genome Sequencing Project.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Strong C., Graves T., Duckels G.;
RT "The sequence of A. thaliana F2P3.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Waterston R.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Bevan M., Pohl T., Weizenegger T., Bancroft I., Mewes H.W.,
  Mayer K.F.X., Schueller C.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Pohl T., Weizenegger T., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
```

DR EMBL; AF080120; AAC35535.1; -  
DR EMBL; AL049876; CAB43048.1; -  
DR EMBL; AL161531; CAB81214.1; -  
KW Nucleotide transferase, RNA-directed RNA polymerase; Transferase.  
SQ SEQUENCE 1133 AA; 149323 MW; 32B72C4E429B20B9 CRC64;  
  
Query Match 34.0%; Score 1995.5; DB 10; Length 1133;  
Best Local Similarity 40.1%; Pred. No. 7e-135;  
Matches 459; Conservative 214; Mismatches 389; Indels 83; Gaps 31;  
  
QY 4 TTQVGFYLLSAEYVVFSELEKYTGCTVCALEVQKSGG--SRAFAKVVQAD-NISADK 60  
DB 11 TVKISNVPTQIYVADSELLRFLHGEDVFALEIPTTRDNMKPRDFARVQVTTLEVKSA 70  
QY 61 IITLANNRLYGGSYLKAWEKMTDIVQLRAYVD--QMDGITLNFQCGISDDK-FAVLGS 116  
DB 71 QLLSSQSKLLFKTHNLRLSEAYDIIP--RPVDPKRLDDIVLVGPFSEDEKRECALEK 128  
QY 117 TEVSTQFGIGLKK---FFFFSSGADYKQLQSLNINWQVVLHRYPYGONAQF--LLIQLF 171  
DB 129 WD-GVRCWILTEKRRVEFWWESGDC-YKIEVRFEDIITLSCCVNGDASEIDAFLLKLK 186  
QY 172 GAPRIYKRL-----ENSCYSFFKETPDDQWVNTDPPPS-WIGLSSSLCLOFRGRVR 222  
DB 187 YGPKYKRVTHIAKFKSDRYCKEDFDWMTITDFTSGSKSIGTSTCFLEWHNGST 246  
QY 223 LPNFESEFFHAERNNITLOTGTFVFSQKSAALVNPVQPP-EGISIPYKILFKISSIVQ 281  
DB 247 MLDIFSGPLFYREDPLSTIYVDGKTF--ASAAQIVPLNAAAILGLEFFYELFOLNVLHV 304  
QY 282 HGCIPGALNVYFRLVDPVRRNV--ACIEHA---LEKLYIKECVDPVWRMLTEQYDG 335  
DB 305 -----AQISLFAASDMELIKILRGMSLETALVILKHLQOQSSICVDPVFFVKTQMS 357  
QY 336 YLK--GROPPKSPSTLDDGLVYVRRVLVTPCKVYFCGPEVNVSNRNLRYNSEDINFLR 393  
DB 358 VYKMKHSPASAYKRLTEQIMSQRAVTPSKYILLGPELETANYVVKNAEHSVDFMR 417  
QY 394 VSFVDEEWEKLYSTDLLPKASTGVS---RNIYERILSTLRKGFVIGDKKFEELARSS 450  
DB 418 VTFVEEDSKVLPANALSVNSKEGYFVKPSRTNIYNRLVILGEGITVPKRFELAFAS 477  
QY 451 QLDRNSVMEFASRPGLTANDIRAMGDFSQIKNVAKYAALRGQFSGSRETLVLRHEIE 510  
DB 478 QLGRNSVMEFASNEKVKAEIDREWNGCFKIRISIKCAARMGQLFSASRQLIYRAQDVE 537  
QY 511 VTPDVKV--HGTSYVFSOGIKSIGDFAHRVASKGLQYTPSAFQIRYGGYKGVGVGDDP 568  
DB 538 QIPDIEVTTDGDYCFDGIKISLAFAKVAQKACGLSHVPSAFQIRYGGYKGVIAVDRS 597  
QY 569 SSKLSLRKSKSYESDNIKLDVLGWSKYQPCYNLRQLITLLSTLGVKDEVLEQKQEA 628  
DB 598 SFRKLSLRDMLKFDNSNRMLNVRTWTSMPCLNREIICLLSLTIGLIEDANFEAMQAVHL 657  
QY 629 DQDAILHDSLKAQPALELMSGENT--NILKAMLNCGYKPDAPFLLSMMLQTFRASKLLD 687  
DB 658 SMLGNMLEDRAALNVQLKLS--GENSKNLLVKMLLQGYASSEPYLSMMLRVHSESQLE 716  
QY 688 LRTSRIFPIPGRTMMGCLDSRTLEYGVQVFTGAGHGEFSDLLHPFNNSRSTNSNFI 747  
DB 717 LKSRCLILVPGRIILIGCDMEGLLEYGVYVRLTLKAEKSKRQSYFRKIDETS--V 774  
QY 748 LKGNVYVAKNCPHLPDGIIRLVKAVNRLHH---WVDCVYVFPQKGRPHPNNECSGSDLD 803  
DB 775 VIGKVVVTKNCPHLPDGIIRVLA--IYEVHFEKGYLDCIIFPQKGERPHPNNECSGGDLD 832  
QY 804 GDIFYVCHDQDMIPRQVQPMYPPAPSIQLDHDVTEVEEYETNYINVDLSLIANAH 863  
DB 833 GDQFFVWDEKLIIPSEMDPPMDYAGSRPLMDHVTLEEHKFFVVDYIMISITLGVISTAH 892  
QY 864 VVFADREPDMASDPCKKLAEFLSIADVFPKTVGTPAEIPSPKPEYDFDKMDKDKTSYI 923  
DB 893 LVHADROPKARSQKCLELANLHSPRAVDFAKTGAPAEKMPYALKPREPDPFLEREKPTIYI 952

QY 924 SERVIGKLFKRVK----DKAPOASSIATFTEDVARRSYDADMEVDCPEDYIDEAFDYKTE 979  
DB 953 SESVFGKLYRAVKSLSLAQRPEASEDTV-----AYDVTLEAGFESFETAKAHRDM 1005  
QY 980 YDNKGLNMDYYGIGKTEAETILSGIMKASKTF---DRRKDA---EASIVAVRALRKEARA 1033  
DB 1006 YGKLTSLMIYVGAANEIEILT-GILTKEMYLARDNRRYGDMDKDRITLSVKDLHKEAMG 1064  
QY 1034 WFKRRNDIDMLPK-ASAWTHVYHPTWYGCYNOGLKRAHFISFPWCYVQQLTQIKKDKA 1092  
DB 1065 WFEKSCDEQCKKLAGAWYVYTNP-----NHRDEKLTFLSPFWIVGVLDLKAENA 1118  
QY 1093 RNRPV 1097  
DB 1119 QRQSV 1123  
  
RESULT 6  
Q92RX5 PRELIMINARY; PRT; 513 AA.  
AC Q92RX5;  
DT 01-MAY-1999 (T-EMBLrel. 10, Created)  
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
DE RNA-directed RNA polymerase (fragment).  
GN RDRP.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA; TISSUE=LEAF;  
RA Schiebel W., Pelissier T., Riedel L., Thalmair S., Schiebel R.,  
RA Kempe D., Lottspeich F., Saenger H.L., Wassenecker M.;  
RT "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from  
RT Tomato";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ011977; CAA09894.1; -  
KW RNA-directed RNA polymerase.  
FT NON\_TER 1 1  
FT NON\_TER 513 513  
SQ SEQUENCE 513 AA; 57802 MW; 924FDA2B70EDA0E CRC64;  
  
Query Match 33.8%; Score 1984; DB 10; Length 513;  
Best Local Similarity 72.9%; Pred. No. 1.3e-134;  
Matches 385; Conservative 50; Mismatches 75; Indels 18; Gaps 4;  
  
QY 471 IRAMWDFSQIKNVAKYAALRGQFSGSRETLVLRHEIEVDPVKV--HGTSYVFSDDGI 528  
DB 1 IRAMWGFDIRNVAKYAALRGQFSGSRETLNVRSDIEIVDFEIIISLGTIFYVSDGI 60  
QY 529 GKISGDFAHRVASKGL-QYTPSAFQIRYGGYKGVGVGDDPSSMKLSLRKSKSYESDNI 587  
DB 61 GKISAEFARVARKCGLTEFSPSAFQIRYGGYKGVAVDPNSSKLSLRKSKSFESNT 120  
QY 588 KLDVLGWSKYQPCYNLRQLITLLSTLGVKDEVLEQKQEAVDQDLDAILHDSLKAQEALE 647  
DB 121 KLDVLAWSKQPCYMNQLITLLSTLGVTSVFEKQREVVDRDLALLTHPLFAHALGL 180  
QY 648 MSPGENTNILKAMLNCGYKPDAPFLLSMMLQTFRASKLLDRLTRSRIFIPNGRTMMGCLD 707  
DB 181 MAPGENTNILKALITCGYKPDAPFLLSMMLQNFASKLLELRKTRIFISGGRSMGCLD 240  
QY 708 ESRILEYGVQVFTGAGHGEFSDLLHPFNNSRSTNSNFIKGNVYVAKNCPHLPDGI 767  
DB 241 ETRTLEYGVVQV-----YSDMPRP-----GRRFITGPVVVAKNCPHLSGDVR 285  
QY 768 LKAVNVNALHMDVCCVFPQKGRPHPNNECSGSDLDGDIYFVCWQDMIPRQVQPMY 827  
DB 286 LQAVNVPALNHMDVCCVFPQKGLRPHNECSGSDLDGDIYFVCWQDELVPRTSEPDY 345



```
RESULT 8
Q9LKP0 PRELIMINARY; PRT; 1196 AA.
AC Q9LKP0;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE RNA-dependent RNA polymerase.
GN SDE1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids ii; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20306669; PubMed=10850496;
RA Dalway T., Hamilton A., Rudd S., Angell S., Baulcombe D.C.;
RT "An RNA-dependent RNA polymerase gene in Arabidopsis is required for
RT posttranscriptional gene silencing mediated by a transgene but not by
RT a virus.";
RL Cell 101:543-553(2000).
KW EMBL; AF268093; AAF74208.1; -.
KW RNA-directed RNA polymerase.
SQ SEQUENCE 1196 AA; 136747 MW; C1B1961CDC11F708 CRC64;

Query Match 27.5%; Score 1610; DB 10; Length 1196;
Best Local Similarity 34.6%; Pred. No. 5.1e-107;
Matches 431; Conservative 188; Mismatches 417; Indels 210; Gaps 42;

QY 5 IQVFGPYLLSAEVAASFLEKTYGYGTCALEVQKSGGSAFAKYOAFADNI-SADKIIT 63
DB 15 VSIGFGESITAKLTLDYLEDEGVIVRCLKTSWTPPGSYNPFETADTSNIPSIDEYKK 74
QY 64 LANNLYFGSSYLKANEMKTDIVQLRAYVD- 97
DB 75 VEPHAFVHFAVFSAGRAMDAOCNLLDQPLKVS LGPKNPSYLNQRRTVTPKLAG 134
QY 98 ITLNFQCOISDDKFAVLGSGTEVSITQGI-----GLKQFFFLSSG-----SADY 141
DB 135 ITLEIGTLVSRDDEFVSWRAE-GVDFLVDPFDNTCKFCPRKSTAFSKDAMHAVINCDY 193
QY 142 KLOUSYNIQOVULHRYPGONAFLLI-QLFGAPRIYKRLNESCYSFFKETP-----DD 194
DB 194 KLELLVRDIQT-----ROYKTLHGFLVLQLASSPRVWYRTADD-DIYDFVPGDLLDDDD 248
QY 195 QWVRTDFPPSWIGLSSSLCLOPRGRVRLNFEESPFHYAERENITLQGFTEFF-----V 250
DB 249 PWIRTTDF--TQVG-ALGRCHSYRLIS-PRYENK-----LRALDYFMRVY 292
QY 251 SQKSALVPN-----VQPPGEGISIPYKILFKISSLVQHGCIIPGPAIIV 292
DB 293 QEERVRWPPRIRNEPCRGEPVSDHFFCIHHKEGIS--FEIMFLVNSVLRGVFNQFOLTE 350
QY 293 YFFRLV--DPRRNVAIEHALEKLYIKCCYD-PYRWLTEQYDGYLK-----GRQ 341
DB 351 RFFDLNRQPKDVIASLXHL-----CTYKRPV-----FDYKRLKLVQEWIQKN 395
QY 342 PPKSPSITLDGLVVRVRLVTPCKVFCGPEVNVSNRVLNYSIEDIDNLRVSVFVDEW 401
DB 396 PKLLGSHQSEDSIEIRLVITPRAVCLPPEVLSNRLVRLRYKAVAEFLRVTFMDESM 455
QY 402 EKLVSITDL-----LPKASTGSGV--RTNIYERILSLRKGFTVGDKKFFELAFSSOLR 453
DB 456 QTINSNVLSYFVAPIVKDLTSSPSQKTYVEKRVKSYLTDGFKLCGRKYSFLAFSANQLR 515
QY 454 DNSVNMFASTRPGLANDIRAMWGDFSQIKNVAKYAALGSGFSRSTLSVLRHEIV-I 512
DB 516 DRSAWFFAEDGKTRVSDIKTWGFEKD-KNVAKCAARMGLCFSTYATVDVMPHEVDEV 574
QY 513 PDVKVHGTSYVFSIGIGKISGDFAHRVASKCGL--QVTPSAFQIRYGYGKGVGVDPDSS 570
DB 575 PDIERNG--YVFSDDIGITITPDLAGVMEKLDLVHVSFPCAYQIRYAGFGVVARWFSKS 632
```

Qy	141	YKLOLSTENTWQVVLHPRYPQCNQAQFTLLIQLGAPRIYKLELNSCYSFFKE-----TPDDQ	195
Db	275	YKLHIKEGVYKELCRTWGEHREQAQAVITLRDPPVAYARDVS--KTFGEURLTWSEND	332
Qy	196	WVRTDTPPPWIGLSSSLCLOFRGGRVLRPNFEESFFHYAERENNITLQTGTFEY--SOK	253
Db	333	WERVVDISP---GLDYS-----KNPVSU-----AENHOYIDLGRWITYWIELDQQ	374
Qy	254	SALV-----PNVQP-----	262
Db	375	STRVMDQVOQYLLDWNLRKTLATVPPELPNQKPKVWDFLDDRYGHDIQQVSSRSWNSDFS	434
Qy	263	---PEGISIPYKILFKISSVLQHGIGPGPALNVYFRLVDRPRRNACIASHALEK--LY	316
Db	435	LLAAPPAISLPDVRVQLEVCISOGIIN-----BHNDREPLE	472
Qy	317	YIKECC-----YDPVRWLTQDGYLKGRO---PPK-----SPSITLDDGLV	355
Db	473	KLMEFCDNDSFKDKRARLILEVADEYAGKRIPOFMELFKDHAALAYFPFSFIPNHCA	532
Qy	356	YVRVLVTPCKVVFQCGPEVNVSNRVRUNYSEDIDNFRVRSFVDEBWE--KLXSTDL-LPKA	413
Db	533	WVRSVTITPRIIVFSTPCVEPTNRVIRQNKHAQDFIRIQFTDEVLEGRIKSGEALP--	590
Qy	414	STGSGVETNIYERILSTLRKGFVIGDKTEFLAFSSQLRDSNVWVWFAISRGCLFANDIRA	473
Db	591	-----LFLAVYRLEKGVAMGPMWKNKFLAFONSQIREAGAFMFCQSNLAKGDMRA	641
Qy	474	WMGDSQIKNVAKYAAARLQSGSSRETLSVLRHETIPIVDVKVHGTYSVFSDIGTKISG	533
Db	642	WMGRFSHIKVIKYYAARLQSGFSTRLVPGIPAPRIVTIPDVEKDG--FCFTDGVGKISP	699
Qy	534	DFAHRAVASKGLOYTSAQIRVGGYKGVGVDPD--SSMKLSLRKSMKSYSDNIKLDVL	592
Db	700	LLAKIVAHDSIDPPSAQFRMGCGKVLVTPDVKGMVEHRIKSYQBEKFAEENGLWEV	759
Qy	593	GSWSYQPCYLNROLITLLSTGLVKDEVLBQKQEAVDQLDALHDSLKAQEALE--LMSPG	651
Db	760	RCSQFSATLNROITIAVLSLSGVPDQVFDVMMEQOLSDFNAAEMDKKATAILKTFIDEN	819
Qy	652	ENTWILKAMLCNGYKPDAPPEFLSMLOTFRASKLLDLRFSRIFIPNGRTWVGCLDESRT	711
Db	820	HMTPIAEMLAYFGMSQSEFFVETLLQLRWSISIKLKEKARLNVEKSAFTVGCVDDEGT	879
Qy	712	LEVQGVVQVTFAGHGEFSDLHPFNNSR-----STNSNF--ILKGNVVAKNPCL	760
Db	880	LK-----GHMKVIEDMKDVSSEKLQIFLIQIPDVGNGYRVITGTTCVVGRRNPSL	928
Qy	761	HPGDIRVLKAVNRALHHMVDCVVPPOKGRPHBPECSSDLGDGIYFVQWQDWIP--PR	819
Db	929	HPGDIRVEAVDPALRHLDVVVPLTGDRVPMSCGGDUGDDFFVWDPDLPIKPER	989
Qy	820	QVQPMYPPAPSIQLDHDVTIEVEEYFTNYIVNDSLGIIANAHVFAADREPDMAMSDPC	879
Db	989	SHPMISEPTAGRELATEFTVNNLTFFVLYMKYNNPLIAHAHLATADAEEVGKSPKC	1048
Qy	880	KKLAEIFSIAVDPKFGVPAETPSOLRKEYPDEFMDKPKDTSYISERVIGLFRKVKDKA	939
Db	1049	LELASTESMAVDYKTVGAABEPRLRDPXTWPHFEK--NRHTYHSVTHALGKLYDMVKRET	1107
Qy	940	-----PQASIAFTTRDVARSYDADMEVDFEYIDEAFDYKTEYDNKJLGNLMDY	990
Db	1108	FDMKENYQLPFDNRILKHTKCRALR-----DG---TLAKARRIKSQYDTAMRRVMCQ	1156
Qy	991	YGIKTAELISGGIMKASTFDRRKDAEAIISVAVRALRKEARFWKRNDDIMDLPKASA	1050
Db	1157	LEIATEEVTWAFVMSKPRVGSYKLDQNVGRESALAKQHPKDOCKKEAG--GDILLSFVSA	1215
Qy	1051	WYHVTHPTVWGCYNOGLKRAH-----FISFPWCYDQLOIIOIKK	1089
Db	1216	MYATVEEYRIALFE--AKQPHVRDGRGLGTGKTIKPTKTPVLSVPFWLFGKGLGAR	1270

RESULT 10					
Q8TGV4	ID	Q8TGV4	PRELIMINARY;	PRT; 1484 AA.	
AC	Q8TGV4;				
DT	01-JUN-2002	(TrEMBLrel. 21, Created)			
DT	01-JUN-2002	(TrEMBLrel. 21, Last sequence update)			
DT	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)			
DE	Putative RNA-dependent RNA polymerase RDP-1.				
GN	RDP-1.				
OS	Diaporphie perijuncta.				
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;				
OC	Diaporthales; Valsaceae; Diaporthe.				
OX	NCBI_TaxID=186170;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
PC	STRAIN-CMW3407;				
RA	Preisig O., Wingfield M.J.;				
RT	Putative fungal RNA-dependent RNA polymerase.";				
RL	Submitted (JAN-2002) to the EMBL/GenBank/DDJB databases.				
DR	EWBL; AF469822; AAL78034.1;				
KW	RNA-directed RNA polymerase.				
SQ	SEQUENCE 1484 AA; 167213 MW; 570AFB8D12547BD CRC64;				
 Query Match 17.68; Score 1031; DB 3; Length 1484; Best Local Similarity 27.99; Pred.No.4.5e-65; Matches 300; Conservative 196; Mismatches 410; Indels 169; Gaps					
Qy	118	EVSIQEGIGLKKEFFFLSGSADYKLGLOSTENWQVYLHRPYGQNAQLLIQLF-GAPRI 176			
Db	300	QFELQTNLGRY-----GCKAFKSQIDFSHTPKTIWGRR--RENGAWALAMLLPSPKL 351			
Qy	177	YKRL-----ENSQSYFKETPDQWVRTTFP-----PSMIGLSSSLCLOFRRGVRLP 224			
Db	352	YGRLPVSHDEQTLNQEW-DLWARQINISEMHKKLRDLPAKLEKQTKWSIEFGRWIT-- 407			
Qy	225	NFESEFFHARENNNTLOTGFTTFVSQKSAALPNVPQPE----GISIPYKIL----- 273			
Db	408	-----YILDMESSLSWEVSRYTLADYNVKVLQVDHFDLVNPGLSNAMKLLDNSAEMT 460			
Qy	274	-----FKISSLVQHCGIPGPALNYVEF-RLV-----DPRRNVACTEHA 311			
Db	461	LELLGSTQSYHLDFPVRYQLEVCIHGSLGLEYGVNAQFLEKLISFEDRAMMLEGVAEA 520			
Qy	312	LEKLYIKKCCYDPVRWLTTEYDGYLKGROPPKSPSITLDGLVYVRVLVTPOKVYFCG 371			
Db	521	NKQFFEPMEFDDP-RIL--HY--WPNARIAPPYA-----TLVRRAVITPTTIYFKT 566			
Qy	372	PEYVNSRVLYRNSDEDINFRLVSFVDE-EWEKLYSTDLLPKASTGSGVRNIVYERILT 430			
Db	567	PCVELFNRLIRKYSLDNRFLRVQFTDEVTFGAIFS-----SDSKNDNLVTRVHV 619			
Qy	431	LKRGFYVGDKKEEFIAFSSSOLDRNSVMWFASRPGULTANDIRAMWDPSQIKNVAKYTAAR 490			
Db	620	MONGVATIGURHKYKFAFNSQFRENAGAFFCEFDTHVTCASIRNMGMGRPPIRSYSGKFAAR 679			
Qy	491	LGSQSSRETLSVLRHIEVIPDVKHVGTSVY--PSDGIGKISGDFAHRVASKCGLOYT 548			
Db	680	MGQCFTTTQVNGISIPKIRIQDIGQRSGDNTWNEDTGVGKISVFYPFARMIASERDLPET 739			
Qy	549	PSAFQTRYGGYKGVGVDDP--SSMKLSLRKSMKEYESDNKLDVLGWSKYOPCVLNRLQI 607			
Db	740	PSCFQMRMGCGKVVVVWPFDPANEVHIRIPSQKFANYNGLEIIKTSTFAHATLNKQVI 799			
Qy	608	TLLSTLGVKDEYLEQKQEAVDQLDALIHDSKAQAELMSFGENTNILKAMILNCQYKP 667			
Db	800	PVLIALGVDDAVFVRMLDDELKEYDEALADSMKAGELLRSQVDENQITTLTWAEVDTFMD 859			
Qy	668	DABPFLSMMLQTFRASKLLDLRTPSRFIPTNGHTMGCLEDISRILEYQVFQFTGAGHG 727			
Db	860	SESFPFLTWILLRWKCWLKRUKQFAISVKNKSAMIFGVVDETGVLR-----GHS 908			
Qy	728	EFSDDIJHPNN-----SRSTNSNFILKGNVYVAKNPCLHPGDIRVLKA 770			



RC	STRAIN-CW 5588;	
RA	Prestis O., Wingfield M.J.;	
RL	Submitted (Nov-2001) to the EMBL/GenBank/DBJ databases.	
RR	EMBL; AF43073; AAL38011.1;	
DR	RNA-directed RNA polymerase.	
KW	NON_TER	
FT	1	
SO	SEQUENCE	1142 AA; 146794 MW; 27D672BA58E0CCBE CRC64;

Query Match 16.8%; Score 987; DB 3; Length 1122;  
Best Local Similarity 30.3%; Pred. No. 4.2e-62;  
Matches 267: Conservative 160; Mismatches 349; Indels 10

QY	268	IPYKILFISSLYOHGICPGPALNVYFFRLV-----DPRRRNVACIEHHALEKLYIYKEOC	322
Db	92	LDFOVRYQLEVCISGHLLEDEYSVNAQFLDKLNSFDADRARMLEGVAEASIQIHPEPMKFI	151
QY	323	YDPVRWLTEQDGYLKGROPPKSPSITLDDGLVYVBRVLVTPCKVYFCGPEVNSVRNLV	382
Db	152	DDP-----KVLHWPNARPPYA-----TLVRRAVITPTTIFYKTCVELTNRIIR	197
QY	383	NYSEIDINFLRVFSVDE-EWEKLYSTDLLFKASTGSGVRTNIYERILSLTRKGFYDKK	441
Db	198	KYSDLNDRLRVQTDIEITFGKIFS-----SODSKDDNLVIRVHRVMQNGISIGDRH	250
QY	442	FEFLAIFSSSLRNSVVMFASRPLGTANDIRAWMGDFSQIKNVAKYAABLQSGFSSRRT	501
Db	251	YRFLLANSOFRENGAFFCOTEDHVHTCDISIREWMGDFRHSVGVKFAATMGQCTTTTRV	310
QY	502	LSVLRIHEIYIPDV---KVHGTSYVFSDGICKISGDFAHRVASKGQLQYTPSAFYRIYGY	559
Db	311	NGISIPNRIKIDDIERTQDGTWNFTDGVGKISNFEAKLIAASERDLPETPSCFQMRIGG	370
QY	560	KGVVGVDPP--SSMKLSLRKSNKSYESDNKILQVLGWSKQPCPYLNRQLITLILSTLGVKDE	618
Db	371	KGVLVWMPDYPPESEVHVRPSQEFKAVYNNLEIKTFAPSHATFNKQVLPVLVALGVDNS	430
QY	619	VLEQKQEAVDQDLAILHSDLSKQAEALMLSPGENTNILKAMLCNCGYKPDAPFPFSLMQL	678
Db	431	VFVSMDDLELREYDEALADSKAGELLRSQVDENQITLMAENVDFMDSKEPFLTWLLR	490
QY	679	TFRASKLLDLRTSRPIPNGRVMGCLDESRTL-----EYQVQVFQT	722
Db	491	LWKCIWLKRLKHFAISVKKKSAMVGVYDIEIGVLRGHSQATEGYNSIESLPQIFLQVP	550
QY	723	GAGHGFSDDLHPFNNSRSTNSFIKLGNNVVAKNCLHPGDIRVLKAVNVALHHMVC	782
Db	551	IEG-----SDGRSTNTEVITGICVGRNLSLHPGDVRYVEADVPELRLKNV	599
QY	783	VVFQPKRPHPEKCSGLDGIYFYCWQDMIPPRQVQPMBYPP-----APSIQLDH--	836
Db	600	VFPKTDGRDIPSCSGDGMDDGYFYWYDERLIPT-----WDHPLNLHDAGSSTLDKPA	655
QY	837	DVTTEEVYEFNTYVNDISI-GITANAHVVPADREPDMAKSDPCKKLAELFIATAVDFPKTG	896
Db	656	DVTIEDYTRFQAQYMKNDISLGRITATAHFAQADQSGGVKNPKCIELAKLHSMADVIKSG	715
QY	897	VPAEIPSQLAPKEYPDMFKPDKTYSISERVIGKLFKRVKDKAPQASSIATF--TRDVARR	955
Db	716	RPVAKRHQLQPRKWPWHMEKKSXYKSYSALCKIYDRIKIEEHAAYEMPPDARILSRY	775
QY	956	SYDADMEVDGFEDYIDAFDYKTEYDNKGLNLDYGIK--TEAEILSGGTMKASKTFDRR	1011
Db	776	QLEADT-----LAKASKIKATYDIAMRLRMGQHEAPVTEFEIWSITFILSRPVRGSDY	827
QY	1015	KDAEASIVAVRALRKEARAWFKR-----RNDIDMDLPKAS-----ANVHTYHP	1051
Db	828	KLOENVGRELAAALKRFAECMEAVTNIQTEGFAPASSVVNLEKLDRFVAAMYTVTHND	887
QY	1059	TYWGCYNQGLKRAH-----FISFPWCYVDQILQI	1087
Db	888	VRAALRERSPKPNGEGSEIEOMPLISFPFLWFHRELARV	927

RESULT 13

```

O14227
ID O14227 PRELIMINARY; PRT; 1215 AA.
AC O14227;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-1998 (TREMBLrel. 06, Last annotation update)
DE Hypothetical 139.5 kDa protein C6F12.09 in chromosome I.
DE SPAC6F12.09.
GN Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OC NCBI_TaxId=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Badcock K., Churcher C.M., Barrell B.G., Rajadream M.A., Wood V.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RC -1- SIMILARITY: SOME, TO C.ELEGANS F10E5.7, F26A3.3 AND
DE P.CHRYSSOSPORUM HYPOTHETICAL 109.5 KDA PROTEIN (Q01869).
CC EMBL; Z98533; CAB11093.1; -
CC Hypothetical protein.
CC SEQUENCE 1215 AA: 389B95C8217CB05C CRC64;
KW

```

Query Match 16.2%; Score 948; DB 3; Length 1215;  
Best Local Similarity 31.3%; Pred. No. 3.2e-59;  
Matches 243; Conservative 155; Mismatches 316; Indels 62; Gaps 18;

[illegible]

```
QY 987 LMDYGIKTEAELSGIM---KASTFDRKDAEIAISVAVRALKEARAWKRRNDIDD 1043
Db 1091 IMARFDISTEYVYTAFLFKDLAKTVNEYGLREEVSFQDLKKKKYQYELERKALSN 1150
QY 1044 MLPKASAWHYHTVH---PTYWCYNGQLKRAH-----FISPPWCYVDQLIQIKK 1089
Db 1151 QSAFDSSEYERINSAVAATYDYDQVKVSGNGITEVLISPYLFSSRLCOLSR 1206

RESULT 14
Q19285
ID Q19285 PRELIMINARY; PRT; 1780 AA.
AC Q19285; Q22323;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (T-EMBLrel. 10, Last annotation update)
DE F10B5.7 protein.
GN F10B5.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Burton J.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Sims M.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z66500; CAA91312.1;
DR EMBL; Z48334; CAA91312.1; JOINED.
DR EMBL; Z48334; CAA8315.1;
DR EMBL; Z66500; CAA8315.1; JOINED.
SQ SEQUENCE 1780 AA; 203145 MW; 16444F29CE9F02C2 CRC64;

Query Match 15.9%; Score 932; DB 5; Length 1780;
Best Local Similarity 25.1%; Pred. No. 8.4e-58;
Matches 322; Conservative 199; Mismatches 465; Indels 296; Gaps 44;

QY 14 LSAEVVSKFLEKYTG-YGTVCALVQKSGKGSRAFAKV--QFAD-NISADKIIFLANRL 69
Db 270 LFSEFASFTSRITGMLHQVFLVPM---HTLFTITPOHMDINISATAIGNCPNSGL 326
QY 70 YFGSSYLKAWEMKTDIVOLRAYVQMDGITLNFQCIQSDDKFAVLGSTE----- 118
Db 327 FLVRGDFISQENTVCSVKLQSH-----HNADASRENSFKVAGSNKYLSYARFEHDK 378
QY 119 --VSTQFGLKKTFFLSSG--SADYKLOLSYENIQVVLHRYPGQNAQFLLLIQLFAP 174
Db 379 RLAVYVFGVRLAE---FADGDHAGFRNLNLYNLFVRIVVDMSH-ETTNSIYIQMNPP 434
QY 175 RIYKRL-ENSCYSFFK-----ETPDQWVRTTDFP-----PSWIGLS 210
Db 435 HLWEGIPKNTIFHRSKSKVLNMT-CIETRVLSWPGDAEGRGVGCTSEAFSSQSWIRLT 493
QY 211 SSLCLOFRGRVRLNFEESFYHAERENNILOTGTFVFSQKSAALVPNPVPPGIGIPY 270
Db 494 MR-----KDDNDNSVSTQLMDIVTRLSA-----RSKA 521
QY 271 KILFKISSILVOHGCIIPGAL-----NYFFRLVDPRRNRVAC----- 307
Db 522 KVMFGSIFSRKKLAPSPAFHSLGSRANYALQALITRGSVFTQLDFDATENIPSSDND 581
QY 308 -----IEHALEKLYY-----IKEC-----CYDPVRWLTE-- 331
Db 582 NDEDDDDVDDTKKQWELVHEPLFLKLVRRGMKESQATEETLEQLLNADFERRQIDVVT 641
QY 332 -----QYDYLKGRQPPKSPSITLDDGL-----VYVRRVLVTPCKVYFCGP 372
Db 642 APTTWYQSRKIQYERLLGES-----LQDVGLAKPLPKNCVSAKVIIVTPSRILLMAP 694
QY 373 EVNVSNRVLNRYNSIDNLFVRSFVDEEWKLYSTDLLPKASTGSGVRTNIYERILSTLR 432
```

```
Db 695 EYVAVNRVVRREGPY--ALRCVFRDDNLRLAIRDF--SINNIDHMSNIVTEGIIYLTK 750
QY 433 KGFVIGDKKFEFLAFSSSQLRDSNVWVFASRPGL-----TANDIRAWMGDFSQIKKVA 485
Db 751 NGIQVADRYVSFLGWSNSQMRQGCYLXAPRVNALTGEVTGVEDIRVWVGDFRAISVP 810
QY 486 KYAARLGQSGFSSRRETLVLRHEIPIVDKVH-----GTSYVFSDDGIG 529
Db 811 KMSRMGQCFTQOPTVYSSVKNIHVENIQVRLRHHHVIPEDEGGVKNKYCFSDGCG 870
QY 530 KISGDFAHRVASKCGLOVTPPSAFQIRYGGYKGVWGVPD-----SSMKLSLRKSKSY-- 582
Db 871 RISIKLATHISKILOLKEVPACQVRFKFGKILVIDPTIDDIINMPKVIYFRKSQKFG 930
QY 583 ---ESDNIKDLVLSKYPQCVNLNQLITLLTLGLGVKDEVLEQK-----OKEAYD 629
Db 931 GGGELODEYLEVYKVPMPSPVCLNRPFIILDOVSEKQSSASHRRITNRVHVYLERELCS 990
QY 630 QLDAILHDSLKAOEALFLMSPGENTNIL-----KAMLNCGYKPDABFFLSMLQTRASKL 685
Db 991 LSNMLINENQAEEV-----NRTNLADWNAASKRAGFELSVDPLIRDMLSIYRYNI 1044
QY 686 LDLRTRSRIFIPN--CRTMVGCLDESRTLEYGVQVFTGAGHGFEESDDLHPFNNSRSTN 743
Db 1045 IHHSKAKIFLPSLGRSMYGVVDETGLLOYGVFIQYS-----PSIRQTS 1090
QY 744 SNEFLK-GNVVYAKNPCLHPGDIRVLKAVNRALHMDVGVFPQKGRPHNECSGSDL 802
Db 1091 NRPILKTGKVLITKNPCHVPGDVVFDVWQPALAHLVDVWVFPQHGPRPHDEMAGSDL 1150
QY 803 DGIYFVCDQDMIPRQVQPMYPPAPSIQLDHDTVIEVEYFTNYIVNDSLGIIANA 862
Db 1151 DGEYSIIMDOEMLLDYNEEAMVFPSSAAEEDKETTDDMVEFFLRYLQDQSIGRMSHA 1210
QY 863 HVYFADREPDMANSDPCKKLAEFLSIADVFPKTPGPAETPSQLRPKEY-PDFMDKPKDTS 921
Db 1211 HLAYADLHG--LFHENCHALAKCAVAVDFPKSGVPAELSSFEQCEMTPDYMMSSGKPM 1268
QY 922 YISERVIGLFRVKVKAQAPASSIATFTRDVARRSYD-----ADMEV-DGFEDYIDEAFD 975
Db 1269 YSTRNLGOLHRRAR-KVEEVLEEFETRGSVFREYDKLICPEDVDVFFGNEIKLVQTLT 1327
QY 976 YKTEYNKLGNDYXGKIETAEILSGIMKASKTFDRRKD-----AEAISVAVRALR 1028
Db 1328 LRDEYVDMQQLDEIGIEDEASVSGHRAASIKRLAGMERDDYSFYHTDKVVELRYEKLY 1387
QY 1029 KEARAWF-----KRRNDIDMLPKASAWHYHTY-HPTY---WGCYNQG 1067
Db 1388 AVFRANKFEFEGGEINIENDGKNTRLKCTKAMHEKIRQWYFVAYVQPKINKAGRCIGQ- 1446
QY 1068 LKRAHPTSPWCYVDQLIQIKK 1089
Db 1447 -----SLPWAVMDALCULRR 1461

RESULT 15
Q9BH56
ID Q9BH56 PRELIMINARY; PRT; 1579 AA.
AC Q9BH56;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE M01G12.12 protein.
GN M01G12.12.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McIay K.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
```



```
[2]
RN SEQUENCE FROM N.A.
RP MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2016(1998).
[3]
RN SEQUENCE FROM N.A.
RA Lennard N.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL132848; CAC35915.1;
DR EMBL: Z81571; CAC35915.1; JOINED.
DR EMBL: Z81571; CAC35914.1;
DR EMBL: AL132848; CAC35914.1; JOINED.
SQ SEQUENCE 1579 AA; 182870 MW; B64517200715E26C CRC64;

Query Match 15.3%; Score 898; DB 5; Length 1579;
Best Local Similarity 25.1%; Pred. No. 2e-55;
Matches 317; Conservative 196; Mismatches 449; Indels 302; Gaps 45;

QY 43 GSRAFAKVOFADNISADKTIITLANNLYFGS-----SYLKAWEMKTDIVQLRA----- 90
Db 101 GSDFWAK-----TLDCEIVEPLAAMYFGNTQGYTIYISHMOISFGKKISADANELLNK 153
QY 91 -----YVDQMDGITLNFQCGISDDKFAVLGSTEVSIOFGIGLKKFFFLSSGSADYKQL 145
Db 154 IVAEFEDRADMTIVTFC-----LKRKRINQITII 184
QY 146 SYENINQVULHRYGQNAQLLQLFGAPRI-----YKRLN-SCYSFF 188
Db 185 RKTIRIIVDQVDMNKTRHIFELACPLIROGSDVDDDKPSTQKPFYKRTNRYSCIG-T 243
QY 189 KETPDDQWRTTDFPSPWIGLSSSLC-----LQPRRGVRL--PNFEE- 228
Db 244 KEYGSPHESAISPSPTIELQKQESDNGSDNDLYRLVLSRLSRTGVQIEFANFPKV 303
QY 229 -----SFFHAERENNITLTGTTFFVFSQKSAVNPVQPEG:SIYPKILFKLISLV 280
Db 304 DVPIGKVPYLYRPTSSKSAFEFYCNCPKPIROGSDVDDDKPSTQKPFYKRTNRYSCIG-T 243
QY 281 QHGCIPI-----GPALNV---YFRLVDPRRNVACIEHALEKLYIKE----- 320
Db 364 SRGAIVKQVLTDEICMQQFLGLITHYL-----ENDKLC EALEDLIYLDGRKRIG 416
QY 321 ---CCYDVP--RWLTEQYDGLKGRPPKSPSITLDDGLYVRRVLVTPCKYVFCGPEVN 375
Db 417 SINKCFHKICQKRLVMQLTNGMSEQE-----IEEGYQVRKVIPTTRVIYTPPEMI 468
QY 376 VSNRVLRYNSEDIDNFRVSVDEWEKLYSTDLILPKASTGSGVRTNIYERILST----- 430
Db 469 MGNRVLNFRDRTGTHVLRVTRDDNNRKM-----RANATGELLIDICVKKY 513
QY 431 LRKGFVIGDKKFEPLAFSSQLRDNVWMPAS-----RPLGTANDIRAW 474
Db 514 LEHGIVVANRDFGLGCGSSQMRDNGAYFWVKNTDNRHKNACKMNSKFKENI--DSVRNQ 571
QY 475 MGDFSQKNVAKYAAARLQSGFGSSRET-LSV-----LRHEIIVDPKVHGTYSYFSDG 527
Db 572 LGNFIQIENIPKMLARLQCGFTQSRLTGVSIGLDPDNYCLTHDLS--GGRSSNGSEYTFSDG 629
QY 528 IGKISGDFAHRVASKGL-OYTPSAFOIRYGVGVGVDP-----DSS 570
Db 630 VGMMSYEFAQEVSLMFGSRVSCFQFRYRGMKGVLADPILDKERNWQEKHGISLSKN 689
QY 571 MKLSLRKSMKSYSDNIKLDVLGWSKYOP---CYLNQPOLITLSTLGKDEVLEQKQKEA 627
Db 690 IKCVFRPSQIKFEKGQILGDQVENVKYSSPVLVALNKPLINIL-----DQVSEMOSLEC 743
QY 628 VQDLDAIHLDSLKAE---ALELMSPGENTNILKAMLN-----CGYKPDABPFLS 674
Db 744 HRRITSRIEELMDLOTLSFAKQMTDEACFRNKLKEFFPRIDIDYLRITNGFTLSNEPFR 803
```

```
QY 675 MWLQOT---FRASKLLDLRTSRIFIPN--GFTMMGCLDESRILEYGOVFOFTGAGHGEF 729
Db 804 SLVKASIKFAITKQL---FKEQIQIPSELGKSLGVVDETGILQYGOVFOIT-KNHRNI 859
QY 730 SDDLHPFNNSRSTNSNFIKGNVYVAKNCPCLHPGDIRVLKAVNVRALHMHVDCVVFQK 789
Db 860 ---LPPRDSNRKVLGSEIVTGTLLTKNPGIVPGDVRIFEAVDIPELHHLCDVVVFPQH 916
QY 790 KRPHNECSGDLDDGDIYFVCDQDMIPPRQVQPMVEYPPAPSIQLDHDVTIEE-VEEYET 848
Db 917 PRPHPEMAGDLDDGDEYSVINDQKLLERNEEAFDAVENKLTQYEWEDIDDLMRDVIY 976
QY 849 NYIVNDSLGIITANAHVVFADPEMDMAMSDPCKKLAELFSIAVDFPKTGVPAE----- 900
Db 977 EYLKDLVGLLANSHLHNSDOYG--LTSRVCMLAKKSCQAVDFSKSGKPPDELQTTWKT 1034
QY 901 -----IPSQIRPKEYPDF-MDKPDKTSYISERVIGKLFKRVKDKAPQASSIATFTRDV 952
Db 1035 DDATGEMIPPE-RAERVDPDYHVGSDHMPKYVSPRLCGKLPREFOG-IDNAMKISSEKSEQ 1092
QY 953 ARRSYDADMEVDGPDYIDEAFDYKTEYDNKLGMLNDYIGIKTEAELISGGIMK-ASKTF 1011
Db 1093 YKIEVDESIRVDGFEYEMEDAKKQLASYNQOLKSTMDTYGIOSEGEIMSGCIIEMNRIS 1152
QY 1012 DRKDD-----AEAISVAVRALRKEARAW-----FKRR 1038
Db 1153 DSDQDDMSFYNTNRMETKMTALVSKERTIFFQOGFQGEVCTLLPDAYNESNFNFECE 1212
QY 1039 NDIDDLMPKASAWYHVYHTYWGYN--OGLKRAHFISFPWCYVDQLIQIKDKKARNRP 1096
Db 1213 NPNEEIRKAVAWYR-----ACYECAKSTREPRKLSFAWIAYDVIAKIKETKVLNV 1264
QY 1097 VLNL 1100
Db 1265 EMNI 1268
```

Search completed: November 6, 2002, 03:46:07  
Job time : 49 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:37:53 ; Search time 167 Seconds  
(without alignments)  
92.517 Million cell updates/sec

Title: US-09-782-874-2  
Perfect score: 1114  
Sequence: 1 MGKTIQVFGPYLLSAEVVK.....RPVLNLSRLAQLSHRLVLK 1114

Scoring table:  
OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 90412 seqs, 13869272 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 90412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database : Published Applications,AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1114	100.0	1114	10	US-09-782-874-2 Sequence 2, Appl
2	218	19.6	218	10	US-09-782-874-3 Sequence 3, Appl
3	17	1.5	17	10	US-09-782-874-12 Sequence 12, Appl
4	14	1.3	14	10	US-09-782-874-11 Sequence 10, Appl
5	14	1.3	14	10	US-09-782-874-11 Sequence 11, Appl
6	10	0.9	13	10	US-09-782-874-13 Sequence 13, Appl
7	8	0.7	283	10	US-09-764-846-220 Sequence 220, App
8	8	0.7	477	10	US-09-764-846-140 Sequence 140, App
9	8	0.7	506	10	US-09-801-368-214 Sequence 214, App
10	7	0.6	18	10	US-09-864-761-45788 Sequence 45788, A
11	7	0.6	24	10	US-09-864-761-46295 Sequence 46295, A
12	7	0.6	46	10	US-09-764-870-336 Sequence 336, App
13	7	0.6	46	10	US-09-764-869-1017 Sequence 1017, Ap
14	7	0.6	51	10	US-09-925-299-1317 Sequence 1317, Ap
15	7	0.6	85	10	US-09-864-761-37705 Sequence 37705, A
16	7	0.6	157	10	US-09-741-569-427 Sequence 427, App
17	7	0.6	249	10	US-09-205-658-309 Sequence 309, App
18	7	0.6	279	10	US-09-789-582A-2 Sequence 2, Appli
19	7	0.6	322	10	US-09-747-835A-54 Sequence 54, Appl

Sequence 11987, A	333	10	US-09-815-242-11987	
Sequence 148, App	358	10	US-09-801-368-148	
Sequence 2, Appli	403	10	US-09-870-379-2	
Sequence 19, Appl	436	8	US-08-980-068B-19	
Sequence 1377, Ap	469	10	US-09-925-301-1377	
Sequence 13863, A	482	10	US-09-815-242-13863	
Sequence 478, Ap	536	10	US-09-815-242-478	
Sequence 10841, A	563	10	US-09-815-242-10841	
Sequence 10240, A	575	10	US-09-815-242-10240	
Sequence 2, Appli	612	8	US-08-910-386A-2	
Sequence 195, App	737	10	US-09-771-161A-195	
Sequence 29, Appl	834	10	US-09-934-909-29	
Sequence 7, Appli	1025	8	US-08-910-386A-7	
Sequence 12, Appl	8	10	US-09-872-349-12	
Sequence 19, Appl	6	0.5	10	US-09-124-280A-19
Sequence 129, App	14	10	US-09-765-527-129	
Sequence 90, Appl	14	10	US-09-881-490-90	
Sequence 101, App	14	10	US-09-881-490-101	
Sequence 43391, A	20	10	US-09-864-761-43391	
Sequence 14, Appl	23	10	US-09-860-793-14	
Sequence 38599, A	25	10	US-09-864-761-38599	
Sequence 13, Appl	26	10	US-09-860-793-13	
Sequence 45443, A	27	10	US-09-864-761-45443	
Sequence 41081, A	32	10	US-09-864-761-41081	
Sequence 47440, A	37	10	US-09-864-761-47440	
Sequence 810, App	38	10	US-09-764-869-810	
Sequence 43869, A	39	10	US-09-864-761-43869	
Sequence 46284, A	39	10	US-09-864-761-46284	
Sequence 36412, A	44	10	US-09-864-761-36412	
Sequence 40514, A	45	10	US-09-864-761-40514	
Sequence 45158, A	45	10	US-09-864-761-45158	

ALIGNMENTS

RESULT 1  
US-09-782-874-2  
; Sequence 2, Application US/09782874  
; Patent No. US20010023067A1  
; GENERAL INFORMATION:  
; APPLICANT: Wassenegeger, Michael  
; Riedel, Leonhard  
; Schiebel, Winfried  
; Sanger, Heinz

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN  
RNA-DIRECTED RNA POLYMERASE (RDRP)

NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSER: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/782,874  
FILING DATE: 08-Feb-2001  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/811,583  
FILING DATE: 05-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley, James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: MPG-1  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-596-9000  
TELEFAX: 212-596-9090  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1114 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-782-874-2

Query Match 100.0%; Score 1114; DB 10; Length 1114;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKTIQVFGPGLLSAEVWSFLEKYGTGTCVCALEVKQSGSRAPAKVQFADNLSADK 60  
DB 1 MGKTIQVFGPGLLSAEVWSFLEKYGTGTCVCALEVKQSGSRAPAKVQFADNLSADK 60

QY 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDQMDGITLFGCQISDDKFAVLGSTEVS 120  
DB 61 IITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDQMDGITLFGCQISDDKFAVLGSTEVS 120

QY 121 IQFGIGLKKFFFLSSGSADYKQLSYENITWVYLHRYPCQNAQFLLIQLFGAPRYKRL 180  
DB 121 IQFGIGLKKFFFLSSGSADYKQLSYENITWVYLHRYPCQNAQFLLIQLFGAPRYKRL 180

QY 181 ENSCYSEFKETPDQWVRTTDFPPSMTGLSSSLCLOPRRGVRLPNPESFFHYAERENNI 240  
DB 181 ENSCYSEFKETPDQWVRTTDFPPSMTGLSSSLCLOPRRGVRLPNPESFFHYAERENNI 240

QY 241 TLQGTFFVFSQKALPNVQPPBGISIPYKILFKISLVOHGCIPGALNVYFFRLVDP 300  
DB 241 TLQGTFFVFSQKALPNVQPPBGISIPYKILFKISLVOHGCIPGALNVYFFRLVDP 300

QY 301 RRRNVACIEHALEKLYIKKCCYDPVWLTEQYDGLKGRQPKPSITLDDGLVYVRRV 360  
DB 301 RRRNVACIEHALEKLYIKKCCYDPVWLTEQYDGLKGRQPKPSITLDDGLVYVRRV 360

QY 361 LVTPCKYVFCGPEVNVNRLNRYSEDIDNLFVSVFDEWEKLYSDLLPKASTGSGVR 420  
DB 361 LVTPCKYVFCGPEVNVNRLNRYSEDIDNLFVSVFDEWEKLYSDLLPKASTGSGVR 420

QY 421 TNYIERILSTLRKGFVIGDKKFFELAPSSQLRDNVWVFASRPGTANDIRAWMGDFSQ 480  
DB 421 TNYIERILSTLRKGFVIGDKKFFELAPSSQLRDNVWVFASRPGTANDIRAWMGDFSQ 480

QY 481 IKNVAKYAARLGQSGFSRSTLSVLRHEIEVDPVKVHGTSYVFSDGIGKISGDFAHRA 540  
DB 481 IKNVAKYAARLGQSGFSRSTLSVLRHEIEVDPVKVHGTSYVFSDGIGKISGDFAHRA 540

QY 541 SKCGLOVTPSAFQIRYGGYGVGVGVDPSMKLSLRKSMKSYSDNKLKDLVLGWSKYQPC 600  
DB 541 SKCGLOVTPSAFQIRYGGYGVGVGVDPSMKLSLRKSMKSYSDNKLKDLVLGWSKYQPC 600

QY 601 YLNQLITLTLSTLGKDEVLEQKQKAVDQDAILHDSLKAQALEMSPGENTNLIKAM 660  
DB 601 YLNQLITLTLSTLGKDEVLEQKQKAVDQDAILHDSLKAQALEMSPGENTNLIKAM 660

QY 661 LNCYKFPDAPFTSMMLQTPRAKLLDLRTRSRIFIPNGRTMGMCLDESSTLYGQVFVQ 720  
DB 661 LNCYKFPDAPFTSMMLQTPRAKLLDLRTRSRIFIPNGRTMGMCLDESSTLYGQVFVQ 720

QY 721 FTGAGHGEFSDLLHPFNNSRSTNSFTLKGNVVAKNPLCPGDIRVLKAVNVALHHMV 780  
DB 721 FTGAGHGEFSDLLHPFNNSRSTNSFTLKGNVVAKNPLCPGDIRVLKAVNVALHHMV 780

QY 781 DCVVFQKGRPHNECSGSDLDGDIVFVCWDQDMIPPROVQPMYPPAPSIQLDHDVTI 840  
DB 781 DCVVFQKGRPHNECSGSDLDGDIVFVCWDQDMIPPROVQPMYPPAPSIQLDHDVTI 840

QY 841 EEVEEYFTNYIVNDSGLIIANAHVVFADREPDNMSDPCKKLAELFSIAVDFFKTVPAE 900

DB 841 EEVEEYFTNYIVNDSGLIIANAHVVFADREPDNMSDPCKKLAELFSIAVDFFKTVPAE 900  
QY 901 IPSQLRPKEYPDMKPKDTSYISERVIGKLFKRYKDKAPQASSIATFTRDVARRSYDAD 960  
DB 901 IPSQLRPKEYPDMKPKDTSYISERVIGKLFKRYKDKAPQASSIATFTRDVARRSYDAD 960  
QY 961 MEVDGFEDYIDEAFDYKTEYDNKLGNDYIGIKTEAEITLGGIMKASKTFDRRKDAEAI 1020  
DB 961 MEVDGFEDYIDEAFDYKTEYDNKLGNDYIGIKTEAEITLGGIMKASKTFDRRKDAEAI 1020

QY 1021 SVAVRALRKEARAFKRRNDIDMLPKASAMHYHTYHTYWCYNQGLKRAHFISFPWCV 1080  
DB 1021 SVAVRALRKEARAFKRRNDIDMLPKASAMHYHTYHTYWCYNQGLKRAHFISFPWCV 1080

QY 1081 YDQLIQIKKDKARNPVLNLSIRAOQLSHRLVVK 1114  
DB 1081 YDQLIQIKKDKARNPVLNLSIRAOQLSHRLVVK 1114

RESULT 2  
US-09-782-874-3  
; Sequence 3, Application US/09782874  
; Patent No. US20010023067A1  
; GENERAL INFORMATION:  
; APPLICANT: Wassenecker, Michael  
; Riedel, Leonhard  
; Schiebel, Winfried  
; Sanger, Heinz  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; POLYPEPTIDES HAVING THE ENZYMIC ACTIVITY OF AN  
; RNA-DIRECTED RNA POLYMERASE (RDRP)  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/09/782,874  
; FILING DATE: 08-Feb-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/811,583  
; FILING DATE: 05-MAR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley, James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: MPG-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 218 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: <Unknown>  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-782-874-3

Query Match 19.6%; Score 218; DB 10; Length 218;  
Best Local Similarity 100.0%; Pred. No. 2,4e-206;  
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 RTMGCLEDSRILEYGVQVFTGAGHGEFSDLLHPFNNSRSTNSFTLKGNVVAKNPC 759

Db 1 RTMGCLESRTLEYGVQVFTGAGHGEFSDDLHPFNNSRSTNSFLKGNVYVAKNPC 60  
QY 760 LHPGDIRVLKAVNRALHHMVDVVFQKGRPHPNESGSDLDGDIYFVCWDQDMIPPR 819  
Db 61 LHPGDIRVLKAVNRALHHMVDVVFQKGRPHPNESGSDLDGDIYFVCWDQDMIPPR 120  
QY 820 QVQPMYPPAPSIQLDHDVTTVEEYFTNTIVNDSLGIIANAHVVPADRPDMAMSDPC 879  
Db 121 QVQPMYPPAPSIQLDHDVTTVEEYFTNTIVNDSLGIIANAHVVPADRPDMAMSDPC 180  
QY 880 KKLAEFSTIADFPTKGVPAIPSPQLRPKEYPDMDKP 917  
Db 181 KKLAEFSTIADFPTKGVPAIPSPQLRPKEYPDMDKP 218  
RESULT 3  
US-09-782-874-12  
; Sequence 12, Application US/09782874  
; Patent No. US20010023067A1  
; GENERAL INFORMATION:  
; APPLICANT: Wassenegger, Michael  
; Riedel, Leonhard  
; Schiebel, Winfried  
; Sanger, Heinz  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN  
; RNA-DIRECTED RNA POLYMERASE (RDRP)  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & NEAVE  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/782,874  
; FILING DATE: 08-Feb-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/811,583  
; FILING DATE: 05-MAR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley, James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: MPG-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: <Unknown>  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
US-09-782-874-12

Query Match 1.5%; Score 17; DB 10; Length 17;  
Best Local Similarity 100.0%; Pred. No. 7.6e-10;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 331 EGYDGYLKGROPPKSPS 347  
Db 1 EGYDGYLKGROPPKSPS 17

RESULT 4  
US-09-782-874-10  
; Sequence 10, Application US/09782874  
; Patent No. US20010023067A1  
; GENERAL INFORMATION:  
; APPLICANT: Wassenegger, Michael  
; Riedel, Leonhard  
; Schiebel, Winfried  
; Sanger, Heinz  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN  
; RNA-DIRECTED RNA POLYMERASE (RDRP)  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & NEAVE  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/782,874  
; FILING DATE: 08-Feb-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/811,583  
; FILING DATE: 05-MAR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley, James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: MPG-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 14 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: <Unknown>  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-09-782-874-10  
Query Match 1.3%; Score 14; DB 10; Length 14;  
Best Local Similarity 100.0%; Pred. No. 5.7e-07;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 377 SNRVLRNYSIEDIDN 390  
Db 1 SNRVLRNYSIEDIDN 14  
RESULT 5  
US-09-782-874-11  
; Sequence 11, Application US/09782874  
; Patent No. US20010023067A1  
; GENERAL INFORMATION:  
; APPLICANT: Wassenegger, Michael  
; Riedel, Leonhard  
; Schiebel, Winfried  
; Sanger, Heinz  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN  
; RNA-DIRECTED RNA POLYMERASE (RDRP)  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:

```
;
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/782,874
; FILING DATE: 08-Feb-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/811,583
; FILING DATE: 05-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-782-874-11
;
; Query Match 1.3%; Score 14; DB 10; Length 14;
; Best Local Similarity 100.0%; Pred. No. 5.7e-07;
; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 1007 ASKTFDRRKDAEAI 1020
; | | | | | | | | | |
; Db 1 ASKTFDRRKDAEAI 14
;
; RESULT 6
; US-09-782-874-13
; Sequence 13, Application US/09782874
; Patent No. US20010023067A1
; GENERAL INFORMATION:
; APPLICANT: Wassensegger, Michael
; Riedel, Leonhard
; Schiebel, Winfried
; Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; RNA-DIRECTED RNA POLYMERASE (RDRP)
;
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/782,874
; FILING DATE: 08-Feb-2001
; CLASSIFICATION: <Unknown>
;
;
;
; ADDRESSSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/782,874
; FILING DATE: 05-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-782-874-13
;
; Query Match 0.9%; Score 10; DB 10; Length 13;
; Best Local Similarity 100.0%; Pred. No. 0.0045;
; Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 784 VFPOKQKGRPH 793
; | | | | | | | | | |
; Db 1 VFPOKQKGRPH 10
;
; RESULT 7
; US-09-764-846-220
; Sequence 220, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT212
; CURRENT APPLICATION NUMBER: US/09/764,846
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 220
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (158)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (252)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (265)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-764-846-220
;
; Query Match 0.7%; Score 8; DB 10; Length 283;
; Best Local Similarity 100.0%; Pred. No. 6;
; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 739 SRSTNSNF 746
; | | | | | | | |
; Db 114 SRSTNSNF 121
;
; RESULT 8
; US-09-764-846-140
; Sequence 140, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
```

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PT212  
 ; CURRENT APPLICATION NUMBER: US/09/764,846  
 ; CURRENT FILING DATE: 2001-01-17  
 ; Prior application data removed - consult PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 348  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 140  
 ; LENGTH: 477  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-764-846-140

Query Match 0.7%; Score 8; DB 10; Length 477;  
 Best Local Similarity 100.0%; Pred. No. 9.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 SRSTNSNF 746  
 Db 114 SRSTNSNF 121

## RESULT 9

US-09-801-368-214  
 ; Sequence 214, Application US/09801368  
 ; Patent No. US20020128250A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Busby, Robert  
 ; APPLICANT: Cali, Brian  
 ; APPLICANT: Hecht, Peter  
 ; APPLICANT: Holtzman, Doug  
 ; APPLICANT: Madden, Kevin  
 ; APPLICANT: Maxon, Mary  
 ; APPLICANT: Milne, Todd  
 ; APPLICANT: No. US20020128250A1man, Thea  
 ; APPLICANT: Royer, John  
 ; APPLICANT: Salama, Sofie  
 ; APPLICANT: Sherman, Amir  
 ; APPLICANT: Silva, Jeff  
 ; APPLICANT: Summers, Eric  
 ; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
 ; FILE REFERENCE: 109272.147  
 ; CURRENT APPLICATION NUMBER: US/09/801,368  
 ; CURRENT FILING DATE: 2001-03-07  
 ; PRIOR APPLICATION NUMBER: US 09/487,558  
 ; PRIOR FILING DATE: 2000-01-19  
 ; PRIOR APPLICATION NUMBER: US 60/160,587  
 ; PRIOR FILING DATE: 1999-10-20  
 ; NUMBER OF SEQ ID NOS: 440  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 214  
 ; LENGTH: 506  
 ; TYPE: PRT  
 ; ORGANISM: Saccharomyces cerevisiae  
 US-09-801-368-214

Query Match 0.7%; Score 8; DB 10; Length 506;  
 Best Local Similarity 100.0%; Pred. No. 9.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 923 ISERVIGK 930  
 Db 314 ISERVIGK 321

## RESULT 10

US-09-864-761-45788  
 ; Sequence 45788, Application US/09864761  
 ; Patent No. US20020048763A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharon G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ; FILE REFERENCE: Acomica-X-1  
 ; CURRENT APPLICATION NUMBER: US/09/864,761  
 ; CURRENT FILING DATE: 2001-05-23  
 ; PRIOR APPLICATION NUMBER: US 60/180,312  
 ; PRIOR FILING DATE: 2000-02-04  
 ; PRIOR APPLICATION NUMBER: US 60/207,456  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: US 09/632,366  
 ; PRIOR FILING DATE: 2000-08-03  
 ; PRIOR APPLICATION NUMBER: GB 24263.6  
 ; PRIOR FILING DATE: 2000-10-04  
 ; PRIOR APPLICATION NUMBER: US 60/236,359  
 ; PRIOR FILING DATE: 2000-09-27  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: US 60/234,687  
 ; PRIOR FILING DATE: 2000-09-21  
 ; PRIOR APPLICATION NUMBER: US 09/608,408  
 ; PRIOR FILING DATE: 2000-06-30  
 ; PRIOR APPLICATION NUMBER: US 09/774,203  
 ; PRIOR FILING DATE: 2001-01-29  
 ; NUMBER OF SEQ ID NOS: 49117  
 ; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 45788  
 ; LENGTH: 18  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO AC022317.4  
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.69  
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.76  
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.67  
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.81  
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.76  
 US-09-864-761-45788

Query Match 0.6%; Score 7; DB 10; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 5.3;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 572 KLSLRKS 578  
 Db 12 KLSLRKS 18

## RESULT 11

US-09-864-761-46295  
 ; Sequence 46295, Application US/09864761  
 ; Patent No. US20020048763A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharon G.  
 ; APPLICANT: Rank, David R.

```
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmca-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 46295
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL138761.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; US-09-864-761-46295

Query Match 0.6%; Score 7; DB 10; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 LSSSLCL 215
Db 7 LSSSLCL 13

RESULT 12
US-09-764-870-336
; Sequence 336, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1317
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
; FILE REFERENCE: PT214
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 336
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (39)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-764-870-336

Query Match 0.6%; Score 7; DB 10; Length 46;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 IGLSSSL 213
Db 28 IGLSSSL 34

RESULT 13
US-09-764-869-1017
; Sequence 1017, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1017
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-764-869-1017

Query Match 0.6%; Score 7; DB 10; Length 46;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 569 SSMKLSL 575
Db 30 SSMKLSL 36

RESULT 14
US-09-925-299-1317
; Sequence 1317, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1317
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
FEATURE:
NAME/KEY: SITE
LOCATION: (2)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (3)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (5)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (14)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (16)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (28)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (35)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (38)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (40)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1317
```

```
Query Match          0.6%; Score 7; DB 10; Length 51;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 128 KKFFFL 134
    |||||
DB 41 KKFFFL 47
```

```
RESULT 15
US-09-864-761-37705
; Sequence 37705, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37705
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF067844.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: SWISSPROT HIT: O08586, EVALUE 2.00e-44
; OTHER INFORMATION: EST_HUMAN HIT: AI222037.1, EVALUE 2.00e-43
US-09-864-761-37705
```

```
Query Match          0.6%; Score 7; DB 10; Length 85;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 539 LKAQAL 645
    |||||
DB 67 LKAQAL 73
```

```
RESULT 16
US-09-741-669-427
; Sequence 427, Application US/09741669
; Patent No. US2002002718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; FILE REFERENCE: ELITRA-009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 427
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-669-427
```

```
Query Match          0.6%; Score 7; DB 10; Length 157;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



QY 604 RQLITLL 610  
 DB 37 ROLITLL 43

```

RESULT 17
US-09-205-658-309
; Sequence 309, Application US/09205658
; Patent No. US20010029617A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Ogg, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLOUCE TOLERANCE CONDITIONS
; FILE REFERENCE: 007867351004
; CURRENT APPLICATION NUMBER: US/09/205.658
; CURRENT FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: 08/857,076
; EARLIER FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: 08/888,534
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: US98/10080
; EARLIER FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 309
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-205-658-309

```

```
Query Match      0.6%; Score 7; DB 10; Length 249;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 639 LKAQEAAL 645  
| | | | |  
Dp 143 LKAQEAAL 149

```

RESULT 18
US-09-789-582A-2
; Sequence 2, Application US/09789582A
; Patent No. US20020137169A1
; GENERAL INFORMATION:
; APPLICANT: BASTUCK, CHRISTINE
; APPLICANT: BATHE, BRIGITTE
; APPLICANT: DUSCH, NICOLE
; APPLICANT: MOCKEL, BETTINA
; APPLICANT: THIERBACH, GEORG
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE NADC GENE AND METHODS OF PRODUCING
; TITLE OF INVENTION: ACID OR NICOTINIC ACID DERIVATIVES
; FILE REFERENCE: 202329USOX
; CURRENT APPLICATION NUMBER: US/09/789,582A
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: DE10055870.4
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatenIn version 3.0
; SEQ ID NO 2
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-789-582A-2

```

```
Query Match          0.6%; Score 7; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 295 FRLVDPR 301  
 |||||  
 Db 56 FRLVDPR 62

RESULT 19  
 US-09-747-835A-54  
 ; Sequence 54, Application US/09747835A  
 ; Patent No. US20020146692A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yamazaki, Victoria  
 ; APPLICANT: Tang, Y. Tom  
 ; APPLICANT: Liu, Chenghua  
 ; APPLICANT: Zhou, Ping  
 ; APPLICANT: Wang, Dunrui  
 ; APPLICANT: Zhang, Jie  
 ; APPLICANT: Ren, Feiyan  
 ; APPLICANT: Asundi, Vinod  
 ; APPLICANT: Drmanac, Radjoe T  
 ; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-LIKE  
 ; TITLE OF INVENTION: LIKE) POLYPEPTIDES AND POLYNUCLEOTIDES  
 ; FILE REFERENCE: HYS-37CIP  
 ; CURRENT APPLICATION NUMBER: US/09/747, 835A  
 ; CURRENT FILING DATE: 2002-03-08  
 ; PRIOR APPLICATION NUMBER: US 09/729,739  
 ; PRIOR FILING DATE: 2000-12-04  
 ; PRIOR APPLICATION NUMBER: US 09/653,450  
 ; PRIOR FILING DATE: 2000-08-31  
 ; PRIOR APPLICATION NUMBER: US 09/620,312  
 ; PRIOR FILING DATE: 2000-07-19  
 ; PRIOR APPLICATION NUMBER: US 09/598,042  
 ; PRIOR FILING DATE: 2000-06-20  
 ; PRIOR APPLICATION NUMBER: US 09/552,317  
 ; PRIOR FILING DATE: 2000-04-25  
 ; PRIOR APPLICATION NUMBER: US 09/488,725  
 ; PRIOR FILING DATE: 2000-01-21  
 ; NUMBER OF SEQ ID NOS: 63  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 54  
 ; LENGTH: 322  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; US-09-747-835A-54

Query Match	0.63;	Score 7;	DB 10;	Length 322;
Best Local Similarity	100.0%;	Pred. No. 64;		
Matches	7;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps

QY 478 FSQIKNV 484  
| | | | |  
Db 228 FSOIKNV 234

RESULT 20  
US-09-815-242-11987  
; Sequence 11987, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carf, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA-011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11987  
; LENGTH: 333  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-11987

Query Match 0.6%; Score 7; DB 10; Length 333;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 832 IQLDHDV 838  
|||||  
Db 115 IQLDHDV 121

RESULT 21  
US-09-801-368-148  
; Sequence 148, Application US/09801368  
; Patent No. US20020128250A1  
; GENERAL INFORMATION:  
; APPLICANT: Busby, Robert  
; APPLICANT: Cali, Brian  
; APPLICANT: Hecht, Peter  
; APPLICANT: Holtzman, Doug  
; APPLICANT: Madden, Kevin  
; APPLICANT: Maxon, Mary  
; APPLICANT: Milne, Todd  
; APPLICANT: No. US20020128250A1man, Thea  
; APPLICANT: Royer, John  
; APPLICANT: Salama, Sofie  
; APPLICANT: Sherman, Amir  
; APPLICANT: Silva, Jeff  
; APPLICANT: Summers, Eric  
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
; FILE REFERENCE: 109272.147  
; CURRENT APPLICATION NUMBER: US/09/801,368  
; PRIOR FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 09/487,558  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: US 60/160,557  
; PRIOR FILING DATE: 1999-10-20  
; NUMBER OF SEQ ID NOS: 440  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 148  
; LENGTH: 358  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-09-801-368-148

Query Match 0.6%; Score 7; DB 10; Length 358;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 QLITLLS 611  
|||||  
Db 347 QLITLLS 353

RESULT 22  
US-09-870-379-2  
; Sequence 2, Application US/09870379  
; Patent No. US20020150954A1

; GENERAL INFORMATION:  
; APPLICANT: Donald L. Durden  
; APPLICANT: ADVANCED RESEARCH & TECHNOLOGY INSTITUTE  
; TITLE OF INVENTION: Compositions and Methods for Identifying  
; FILE REFERENCE: ARTI 0024-US  
; CURRENT APPLICATION NUMBER: US/09/870,379  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: PCT/US01/17358  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: 60/274/167  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/208,437  
; PRIOR FILING DATE: 2000-05-30  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 403  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-870-379-2

Query Match 0.6%; Score 7; DB 10; Length 403;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 639 LKAEAL 645  
|||||  
Db 146 LKAEAL 152

RESULT 23  
US-08-980-068B-19  
; Sequence 19, Application US/08980068B  
; Patent No. US20020081718A1  
; GENERAL INFORMATION:  
; APPLICANT: HOTTA, Yoshiaki  
; TITLE OF INVENTION: A POLYPEPTIDE COMMON TO GLIAL CELLS MISSING (GCM)  
; FILE REFERENCE: 97-1513\*/LC/00653  
; CURRENT APPLICATION NUMBER: US/08/980,068B  
; CURRENT FILING DATE: 1997-11-26  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 436  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-08-980-068B-19

Query Match 0.6%; Score 7; DB 8; Length 436;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 898 PAEIPSQ 904  
|||||  
Db 193 PAEIPSQ 199

RESULT 24  
US-09-925-301-1377  
; Sequence 1377, Application US/09925301  
; Patent No. US20020052308A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA106  
; CURRENT APPLICATION NUMBER: US/09/925,301  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05882  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12

```
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1377
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-301-1377

Query Match      0.6%; Score 7; DB 10; Length 469;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 SFFKTEP 192
      |||||
Db 225 SFFKTEP 231

RESULT 25
US-09-815-242-13863
; Sequence 13863, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13863
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-4978

Query Match      0.6%; Score 7; DB 10; Length 536;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1030 ERAWFK 1036
      |||||
Db 183 ERAWFK 189

RESULT 27
US-09-815-242-10841
; Sequence 10841, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13863
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Salmonella typhi
; NAME/KEY: VARIANT
; LOCATION: (1)...(482)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13863

Query Match      0.6%; Score 7; DB 10; Length 482;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1024 VRLRKE 1030
      |||||
Db 399 VRLRKE 405

RESULT 26
US-09-815-242-4978
; Sequence 4978, Application US/09815242
```

; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10841  
; LENGTH: 563  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-815-242-10841

Query Match 0.68; Score 7; DB 10; Length 563;  
Best Local Similarity 100.0%; Pred. No. 1.le-02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1030 ERAWFK 1036  
|||||||  
Db 210 ERAWFK 216

RESULT 28  
US-09-815-242-10240  
; Sequence 10240, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011a  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/181,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10240  
; LENGTH: 575  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-815-242-10240

Query Match 0.68; Score 7; DB 10; Length 575;  
Best Local Similarity 100.0%; Pred. No. 1.le-02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1024 VRLRKE 1030  
|||||||  
Db 399 VRLRKE 405

RESULT 29  
US-08-910-386A-2

; Sequence 2, Application US/08910386A  
; Patent No. US20020092041A1  
; GENERAL INFORMATION:  
; APPLICANT: Ronald, Pamela C.  
; APPLICANT: Wang, Guo-Liang  
; APPLICANT: Song, Wen-Yuang  
; APPLICANT: Hulbert, Scott  
; APPLICANT: Richter, Todd  
; TITLE OF INVENTION: Procedures and Materials for Conferring  
; TITLE OF INVENTION: Disease Resistance in Plants  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/910,386A  
; FILING DATE: 13-AUG-1997  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774  
; REFERENCE/DOCKET NUMBER: 023070-05895005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 612 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-910-386A-2  
  
Query Match 0.68; Score 7; DB 8; Length 612;  
Best Local Similarity 100.0%; Pred. No. 1.le-02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1099 NLSSLRA 1105  
|||||||  
Db 246 NLSSLRA 252  
  
RESULT 30  
US-09-771-161A-195  
; Sequence 195, Application US/09771161A  
; Patent No. US20020110811A1  
; GENERAL INFORMATION:  
; APPLICANT: LEVINE, et al.  
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES  
; FILE REFERENCE: 802620-2005.1  
; CURRENT APPLICATION NUMBER: US/09/771,161A  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: 09/724,676  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: 136776  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: 135619  
; PRIOR FILING DATE: 2000-04-12  
; NUMBER OF SEQ ID NOS: 273  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 195  
; LENGTH: 737  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-09-771-161A-195

Query Match 0.6%; Score 7; DB 10; Length 737;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 647 LMSPGEN 653  
Db 386 LMSPGEN 392

RESULT 31

US-09-934-909-29  
Sequence 29, Application US/09934909  
Patent No. US20020058275A1  
GENERAL INFORMATION:  
APPLICANT: FISHEL, Richard A.  
APPLICANT: GRADIA, Scott  
APPLICANT: ACHARYA, Samir  
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR EFFECTING ADENINE  
TITLE OF INVENTION: NUCLEOTIDE MODULATION OF DNA MISMATCH RECOGNITION  
TITLE OF INVENTION: PROTEINS  
FILE REFERENCE: 9855-6U3  
CURRENT APPLICATION NUMBER: US/09/934,909  
CURRENT FILING DATE: 2001-08-22  
PRIOR APPLICATION NUMBER: US 60/093,935  
PRIOR FILING DATE: 1998-07-23  
PRIOR APPLICATION NUMBER: US 60/066,977  
PRIOR FILING DATE: 1997-11-28  
PRIOR APPLICATION NUMBER: US 60/057,136  
PRIOR FILING DATE: 1997-08-28  
PRIOR APPLICATION NUMBER: US 09/143,571  
PRIOR FILING DATE: 1998-08-28  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 29  
LENGTH: 834  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-934-909-29

Query Match 0.6%; Score 7; DB 10; Length 834;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 728 EFSDDLH 734  
Db 366 EFSDDLH 372

RESULT 32

US-08-910-386A-7  
Sequence 7, Application US/08910386A  
Patent No. US20020092041A1  
GENERAL INFORMATION:  
APPLICANT: Ronald, Pamela C.  
APPLICANT: Wang, Guo-Liang  
APPLICANT: Song, Wen-Yuang  
APPLICANT: Hulbert, Scott  
APPLICANT: Richter, Todd  
TITLE OF INVENTION: Procedures and Materials for Conferring  
TITLE OF INVENTION: Disease Resistance in Plants  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910,386A  
FILING DATE: 13-AUG-1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 023070-0589500S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1025 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-910-386A-7

Query Match 0.6%; Score 7; DB 8; Length 1025;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1099 NLSSLRA 1105  
Db 246 NLSSLRA 252

RESULT 33

US-09-872-349-12  
Sequence 12, Application US/09872349  
Patent No. US20020132980A1  
GENERAL INFORMATION:  
APPLICANT: Sidney Pestka  
TITLE OF INVENTION: PHOSPHORYLATED PROTEINS AND USES RELATED THERETO  
FILE REFERENCE: PBLI-P01-007  
CURRENT APPLICATION NUMBER: US/09/872,349  
CURRENT FILING DATE: 2001-05-31  
PRIOR APPLICATION NUMBER: 60/208,240  
PRIOR FILING DATE: 2000-05-31  
PRIOR APPLICATION NUMBER: 60/255,296  
PRIOR FILING DATE: 2000-12-13  
NUMBER OF SEQ ID NOS: 46  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 12  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: phosphorylated  
OTHER INFORMATION: peptide  
US-09-872-349-12

Query Match 0.5%; Score 6; DB 10; Length 8;  
Best Local Similarity 100.0%; Pred. No. 7.5e+04;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1100 LSSLRA 1105  
Db 3 LSSLRA 8

RESULT 34

US-09-124-280A-19  
Sequence 19, Application US/09124280A  
Patent No. US20020034520A1  
GENERAL INFORMATION:  
APPLICANT: Porro, Massimo  
TITLE OF INVENTION: VACCINES FOR PREVENTION OF GRAM-  
TITLE OF INVENTION: NEGATIVE BACTERIAL INFECTIONS AND ENDOTOXIN RELATED DISEASE  
NUMBER OF SEQUENCES: 45

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Hedman, Gibson & Costigan
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/124,280A
FILING DATE: July 29,1998
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Costigan, James V.
REGISTRATION NUMBER: 25,669
REFERENCE/DOCKET NUMBER: 576-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 302-8989
TELEFAX: (212) 302-8998
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
TOPOLOGY: circular
US-09-124-280A-19

Query Match          0.5%; Score 6; DB 10; Length 10;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 KRFFFF 133
DB 4 KRFFFF 9

RESULT 35
US-09-765-527-129
Sequence 129, Application US/09765527
Patent No. US2002006638A1
GENERAL INFORMATION:
APPLICANT: Better, Marc D.
TITLE OF INVENTION: Methods for Recombinant Microbial Production of
Fusion Proteins and BPI-Derived Peptides
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,527
FILING DATE: 18-Jan-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/621,803
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25,447
REFERENCE/DOCKET NUMBER: 27129/33199

CORRESPONDENCE ADDRESS:
ADDRESSEE: Hedman, Gibson & Costigan
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/124,280A
FILING DATE: July 29,1998
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Costigan, James V.
REGISTRATION NUMBER: 25,669
REFERENCE/DOCKET NUMBER: 576-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 302-8989
TELEFAX: (212) 302-8998
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
TOPOLOGY: circular
US-09-124-280A-19

Query Match          0.5%; Score 6; DB 10; Length 10;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 KRFFFF 133
DB 4 KRFFFF 9

RESULT 36
US-09-881-490-90
Sequence 90, Application US/09881490
Patent No. US20020077298A1
GENERAL INFORMATION:
APPLICANT: Little II, Roger G.
Lim, Edward
Fadem, Mitchell B.
TITLE OF INVENTION: Anti-Fungal Peptides
NUMBER OF SEQUENCES: 211
CORRESPONDENCE ADDRESS:
ADDRESSEE: McAndrews, Held & Malloy, Ltd.
STREET: 500 West Madison Street, 34th FloorDrive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60661
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/881,490
FILING DATE: 14-Jun-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/119,858
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/372,105
FILING DATE: 13-JAN-95
APPLICATION NUMBER: 08/306,473
FILING DATE: 15-SEP-94
APPLICATION NUMBER: 08/273,540
FILING DATE: 11-JUL-94
APPLICATION NUMBER: 08/209,762
FILING DATE: 11-MAR-94
APPLICATION NUMBER: 08/183,222
FILING DATE: 14-JAN-94
APPLICATION NUMBER: 08/093,202
FILING DATE: 15-JUL-93
APPLICATION NUMBER: 08/030,644
FILING DATE: 12-MAR-93
ATTORNEY/AGENT INFORMATION:
```

NAME: McNicholas, Janet M.  
REGISTRATION NUMBER: 32,918  
REFERENCE/DOCKET NUMBER: 100-238/11021US01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/707-8889  
TELEFAX: 312/707-9155  
TELEX: 650 388-1248  
INFORMATION FOR SEQ ID NO: 90:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: misc.feature  
OTHER INFORMATION: "XMP.250"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: C-Terminus  
OTHER INFORMATION: /label= Amidation  
/note= "The C-Terminus is Amidated"  
SEQUENCE DESCRIPTION: SEQ ID NO: 90:  
US-09-881-490-90  
Query Match 0.5%; Score 6; DB 10; Length 14;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 166 LLIQLF 171  
Db 6 LLIQLF 11  
RESULT 37  
US-09-881-490-101  
Sequence 101, Application US/09881490  
Patent No. US20020077298A1  
GENERAL INFORMATION:  
APPLICANT: Little II, Roger G.  
Lim, Edward  
Fadem, Mitchell B.  
TITLE OF INVENTION: Anti-Fungal Peptides  
NUMBER OF SEQUENCES: 211  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McAndrews, Held & Malloy, Ltd.  
STREET: 500 West Madison Street, 34th FloorDrive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60661  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/881,490  
FILING DATE: 14-Jun-2001  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/119,858  
FILING DATE: <Unknown>  
APPLICATION NUMBER: 08/372,105  
FILING DATE: 13-JAN-95  
APPLICATION NUMBER: 08/306,473  
FILING DATE: 15-SEP-94  
APPLICATION NUMBER: 08/273,540  
FILING DATE: 11-JUL-94  
APPLICATION NUMBER: 08/209,762  
FILING DATE: 11-MAR-94  
APPLICATION NUMBER: 08/183,222  
FILING DATE: 14-JAN-94  
APPLICATION NUMBER: 08/093,202  
FILING DATE: 15-JUL-93

APPLICATION NUMBER: 08/030,644  
FILING DATE: 12-MAR-93  
ATTORNEY/AGENT INFORMATION:  
NAME: McNicholas, Janet M.  
REGISTRATION NUMBER: 32,918  
REFERENCE/DOCKET NUMBER: 100-238/11021US01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/707-8889  
TELEFAX: 312/707-9155  
TELEX: 650 388-1248  
INFORMATION FOR SEQ ID NO: 101:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: misc.feature  
OTHER INFORMATION: "XMP.261"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 6  
OTHER INFORMATION: /label= N-methyl-L  
/note= "Position 6 is N-Methyl-leucine"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: C-Terminus  
OTHER INFORMATION: /label= Amidation  
/note= "The C-Terminus is Amidated"  
SEQUENCE DESCRIPTION: SEQ ID NO: 101:  
US-09-881-490-101  
Query Match 0.5%; Score 6; DB 10; Length 14;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 166 LLIQLF 171  
Db 6 LLIQLF 11  
RESULT 38  
US-09-864-761-43391  
Sequence 43391, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
FILE REFERENCE: Aeonica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
SEQ ID NO 43391  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC006385.3  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.85  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.83  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.9  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.98  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.77  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.84  
US-09-864-761-43391

Query Match 0.5%; Score 6; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 902 PSQLRP 907  
DB 6 PSQLRP 11  
|||||

RESULT 39  
US-09-860-793-14  
Sequence 14, Application US/09860793  
Patent No. US20020136734A1  
GENERAL INFORMATION:  
APPLICANT: Pruett, John H  
APPLICANT: Temeyer, Kevin B  
APPLICANT: Kunz, Sidney E  
APPLICANT: Fisher, William F  
TITLE OF INVENTION: Vaccines for the Protection of Cattle from Psoroptic  
FILE REFERENCE: Docket 0047.96 - John H. Pruett et al.  
CURRENT APPLICATION NUMBER: US/09/860.793  
CURRENT FILING DATE: 2001-05-18  
PRIOR APPLICATION NUMBER: 09/366,603  
PRIOR FILING DATE: 1999-08-03  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn ver. 2.1  
SEQ ID NO 14  
LENGTH: 23  
TYPE: PRT  
ORGANISM: Psoroptes ovis  
US-09-860-793-14

Query Match 0.5%; Score 6; DB 10; Length 23;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 YTGyGT 31  
|||||

DB 9 YTGyGT 14

## RESULT 40

US-09-864-761-38599  
Sequence 38599, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
SEQ ID NO 38599  
LENGTH: 25  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO Z83844.5  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9  
OTHER INFORMATION: SWISSPROT HIT: P55194, EVALUATE 2.00e-01  
US-09-864-761-38599

Query Match 0.5%; Score 6; DB 10; Length 25;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DPDSM 571  
|||||



Db 20 DPDSM 25

RESULT 41

US-09-860-793-13  
; Sequence 13, Application US/09860793  
; Patent No. US20020136734A1  
; GENERAL INFORMATION:  
; APPLICANT: Pruett, John H  
; APPLICANT: Temeyer, Kevin B  
; APPLICANT: Kunz, Sidney E  
; APPLICANT: Fisher, William F  
; TITLE OF INVENTION: Vaccines for the Protection of Cattle from Psoroptic  
; FILE REFERENCE: Docket 0047.96 - John H. Pruett et al.  
; CURRENT APPLICATION NUMBER: US/09/860,793  
; CURRENT FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/366,603  
; PRIOR FILING DATE: 1999-08-03  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 26  
; TYPE: PRT  
; ORGANISM: Psoroptes ovis  
US-09-860-793-13

Query Match 0.5%; Score 6; DB 10; Length 26;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 VTGYGT 31

Db 21 YIGYGT 26

RESULT 42

US-09-864-761-45443  
; Sequence 45443, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 60/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 45443  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC015897.3  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.94  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.82  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.94  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.84  
; OTHER INFORMATION: EST\_HUMAN HIT: BF362607.1, EVALUE 1.10e-01  
US-09-864-761-45443

Query Match 0.5%; Score 6; DB 10; Length 27;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1025 RALRKE 1030

Db 20 RALRKE 25

RESULT 43

US-09-864-761-41081  
; Sequence 41081, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 41081  
; LENGTH: 32  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC021677.3  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.8  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.7  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.7  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.8  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3  
; US-09-864-761-41081

Query Match 0.5%; Score 6; DB 10; Length 32;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 FGPPYL 13  
|||||  
Db 18 FGPPYL 23

RESULT 44  
US-09-864-761-47440  
; Sequence 47440, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aeonica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 47440  
; LENGTH: 37  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL132857.3  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.51  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.52  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.44  
; OTHER INFORMATION: SWISSPROT HIT: P77891, EVALUE 4.90e+00  
; US-09-864-761-47440

Query Match 0.5%; Score 6; DB 10; Length 37;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 379 RVLRYN 384  
|||||  
Db 21 RVLRYN 26

RESULT 45  
US-09-764-869-810  
; Sequence 810, Application US/09764869  
; Patent No. US20020061521A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC007  
; CURRENT APPLICATION NUMBER: US/09/764,869  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 2442  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 810  
; LENGTH: 38  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-764-869-810

Query Match 0.5%; Score 6; DB 10; Length 38;  
Best Local Similarity 100.0%; Pred. No. 97;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 ETLSVL 505  
|||||  
Db 26 ETLSVL 31

RESULT 46  
US-09-864-761-45869

; Sequence 45869, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharon G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

; FILE REFERENCE: Aeonica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117

; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1

; SEQ ID NO 45869

; LENGTH: 39

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: MAP TO AC016948.4

; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2

; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1

; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3

; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2

; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1

; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93

; OTHER INFORMATION: EST\_HUMAN HIT: BE880522.1, EVALUATE 7.20e+00

US-09-864-761-45869

Query Match 0.5%; Score 6; DB 10; Length 39;

Best Local Similarity 100.0%; Pred. No. 99;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 502 LSVLRH 507

Db 19 LSVLRH 24

|||||

RESULT 47

US-09-864-761-46284

; Sequence 46284, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharon G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO

; FILE REFERENCE: Aeonica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; NUMBER OF SEQ ID NOS: 49117

; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1

; SEQ ID NO 46284

; LENGTH: 39

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: MAP TO AC008045.5

; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6

; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4

; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.3

; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.4

; OTHER INFORMATION: EST\_HUMAN HIT: BE348399.1, EVALUATE 9.00e-07

; OTHER INFORMATION: SWISSPROT HIT: P55200, EVALUATE 4.30e-01

US-09-864-761-46284

Query Match 0.5%; Score 6; DB 10; Length 39;

Best Local Similarity 100.0%; Pred. No. 99;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 688 LRTRSR 693

Db 19 LRTRSR 24

|||||

DB 2 LRTRSR 7

US-09-864-761-36412

Query Match 0.5%; Score 6; DB 10; Length 44;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 688 LRTRSR 693  
DB 7 LRTRSR 12

RESULT 49

US-09-864-761-40514  
; Sequence 36412, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 36412  
; LENGTH: 44  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC008045.4  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.7  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4  
; OTHER INFORMATION: EXPRESSED IN HSL100, SIGNAL = 1.5  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5  
; OTHER INFORMATION: SWISSPROT HIT: P55200, EVALUE 3.10e-01  
; OTHER INFORMATION: EST\_HUMAN HIT: AW135326.1, EVALUE 9.00e-09

US-09-864-761-36412

Query Match 0.5%; Score 6; DB 10; Length 44;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 688 LRTRSR 693  
DB 7 LRTRSR 12

RESULT 48

US-09-864-761-36412  
; Sequence 36412, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 36412  
; LENGTH: 44  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC008045.4  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.7  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4  
; OTHER INFORMATION: EXPRESSED IN HSL100, SIGNAL = 1.5  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5  
; OTHER INFORMATION: SWISSPROT HIT: P55200, EVALUE 3.10e-01  
; OTHER INFORMATION: EST\_HUMAN HIT: AW135326.1, EVALUE 9.00e-09

OTHER INFORMATION: MAP TO AL137853.3  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.56  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.73  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.46  
OTHER INFORMATION: EST\_HUMAN HIT: BE733841.1, EVALUE 4.00e+00  
US-09-864-761-45158

Query Match 0.5%; Score 6; DB 10; Length 45;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 902 PSQLRP 907  
Db 17 PSQLRP 22

Search completed: November 6, 2002, 03:43:47  
Job time : 175 secs

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.2  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.5  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.9  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2  
OTHER INFORMATION: EST\_HUMAN HIT: BF061077.1, EVALUE 1.00e-10  
US-09-864-761-40514

Query Match 0.5%; Score 6; DB 10; Length 45;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 LSAEVV 19  
Db 29 LSAEVV 34

RESULT 50

US-09-864-761-45158  
Sequence 45158, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecomica-x-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 45158  
LENGTH: 45  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:

ORGANISM: Homo sapiens  
FEATURE:

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:35:08 ; Search time 81 Seconds  
(without alignments)  
1322.145 Million cell updates/sec

Title: US-09-782-874-2  
Perfect score: 1114  
Sequence: 1 MGKTIQVGFPPYLLSAEVK.....RPVLNLSLRAQLSHRLVLK 1114

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: listing first 50 summaries

Database : PIR73:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1114	100.0	1114	2 T30819	RNA-directed RNA p
2	42	3.8	1116	2 T30828	RNA-directed RNA p
3	14	1.3	1133	2 T01920	probable RNA-direc
4	9	0.8	583	2 H69165	hypothetical prote
5	8	0.7	230	2 T22763	hypothetical prote
6	8	0.7	265	2 S60947	hypothetical prote
7	8	0.7	286	2 A95395	protein (imported
8	8	0.7	341	2 T38874	probable short cha
9	8	0.7	352	2 H37272	histidinol-phospha
10	8	0.7	386	2 S21436	probable membrane
11	8	0.7	506	2 S69045	protein kinase MKK
12	8	0.7	641	2 G28771	hypothetical prote
13	8	0.7	830	2 T04848	protein kinase hom
14	8	0.7	901	2 D70116	transcription elon
15	8	0.7	1101	2 G70951	probable ATP-depen
16	8	0.7	1385	2 S42330	156k protein - Pla
17	8	0.7	1564	2 T27121	hypothetical prote
18	7	0.6	29	2 AB0717	hypothetical prote
19	7	0.6	48	2 A57125	polyphenolic adhes
20	7	0.6	66	2 AB2371	hypothetical prote
21	7	0.6	70	2 S68954	polyphenolic adhes
22	7	0.6	79	2 A01030	probable membrane
23	7	0.6	89	2 T46384	hypothetical prote
24	7	0.6	91	2 T07231	hypothetical prote
25	7	0.6	112	2 S14355	glutathione trans
26	7	0.6	116	2 AG2329	50S ribosomal prot
27	7	0.6	121	2 D97840	hypothetical prote
28	7	0.6	122	2 AC2172	hypothetical prote
29	7	0.6	124	2 AG1754	bacteriophage prot

30	7	0.6	141	2 H71504	ribosomal protein
31	7	0.6	142	2 A81663	ribosomal protein
32	7	0.6	145	2 H82567	outer membrane pro
33	7	0.6	149	2 B89329	hypothetical prote
34	7	0.6	151	2 A60943	Ig heavy chain pre
35	7	0.6	154	2 S75019	glutathione peroxi
36	7	0.6	157	2 H64930	hypothetical prote
37	7	0.6	160	2 F83541	probable glutathio
38	7	0.6	167	2 T14262	glutathione peroxi
39	7	0.6	169	2 D84722	probable glutathio
40	7	0.6	183	1 QRECEB	vitamin B12 transp
41	7	0.6	183	2 A98931	vitamin B12 transp
42	7	0.6	183	2 E85779	vitamin B12 transp
43	7	0.6	183	2 S04743	TpA-induced protei
44	7	0.6	197	2 T15923	hypothetical prote
45	7	0.6	206	2 A84865	probable glutathio
46	7	0.6	206	2 E75310	hypothetical prote
47	7	0.6	207	2 F83815	hypothetical prote
48	7	0.6	211	1 JC2368	ribosomal protein
49	7	0.6	211	2 S23753	ribosomal protein
50	7	0.6	213	2 T22984	hypothetical prote

ALIGNMENTS

RESULT 1

T30819  
RNA-directed RNA polymerase (EC 2.7.7.48) - tomato  
C:Species: Lycopersicon esculentum (tomato)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 21-Jul-2000  
C:Accession: T30819  
R:Schiebel, W.; Fellssier, T.; Riedel, L.; Thalmair, S.; Schiebel, R.; Kempe, D.; Lo  
Plant Cell 10, 2087-2102, 1998  
A:Title: Isolation of an RNA-directed RNA polymerase-specific cDNA clone from tomato.  
A:Reference number: Z20885; MUID:99055198; PMID:9836747  
A:Accession: T30819  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1114 <SCH>  
A:Cross-references: EMBL:Y10403; NID:g4038591; PIDN:CAA71421.1; PID:g4038592  
C:Genetics:  
A:Note: RdRP  
C:Superfamily: Arabidopsis probable RNA-directed RNA polymerase  
C:Keywords: nucleotidyltransferase

Query Match 100.0%; Score 1114; DB 2; Length 1114;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MGKTIQVGFPPYLLSAEVKSFLEKTYGTVCALEVKQSKGGSRAFAKVFADNISADK	60
Db	1	MGKTIQVGFPPYLLSAEVKSFLEKTYGTVCALEVKQSKGGSRAFAKVFADNISADK	60
Qy	61	IITLANNRLYPGSSYLKAWEMKTDIVQLRAYVDMQDITLNFQGISDDKFAVLGSFEVS	120
Db	61	IITLANNRLYPGSSYLKAWEMKTDIVQLRAYVDMQDITLNFQGISDDKFAVLGSFEVS	120
Qy	121	IQFGIGLKPPFFLLSSGSADYKQLSYENIWVYVHRPYGNAQFLLIQFGAPRIYKRL	180
Db	121	IQFGIGLKPPFFLLSSGSADYKQLSYENIWVYVHRPYGNAQFLLIQFGAPRIYKRL	180
Qy	181	ENSCYSFFKETPDQWVTTDFPSSWIGLSSSLCQFRGVRVLPNFESFFHYAERENNI	240
Db	181	ENSCYSFFKETPDQWVTTDFPSSWIGLSSSLCQFRGVRVLPNFESFFHYAERENNI	240
Qy	241	TLOGTGFFVFSQKALVPNVOPPEGISIPYKILFKISLVHGGICPGALNVYFRLVDP	300
Db	241	TLOGTGFFVFSQKALVPNVOPPEGISIPYKILFKISLVHGGICPGALNVYFRLVDP	300
Qy	301	RRNVACIEHALEKLYIKKCYDPRWLTEQYDGYLKGROPPKSPSITLDDGLVYVRV	360
Db	301	RRNVACIEHALEKLYIKKCYDPRWLTEQYDGYLKGROPPKSPSITLDDGLVYVRV	360

QY 361 LVTCKVYFCGPEVNVNRLVNYSEDIDNLRVSFVDEWEKLYSTDLKPKASTGSGVR 420  
Db 361 LVTCKVYFCGPEVNVNRLVNYSEDIDNLRVSFVDEWEKLYSTDLKPKASTGSGVR 420  
QY 421 TNYIERILSTLRKGVIGDKFEFLAFSSQLRDNVWVFASRGLTANDIRAWMGDPQS 480  
Db 421 TNYIERILSTLRKGVIGDKFEFLAFSSQLRDNVWVFASRGLTANDIRAWMGDPQS 480  
QY 481 INKVAKYARLQSGFSGSRETLVSRHEIEVDPVKVHGTSYVFSDDGKISGDFAHRYA 540  
Db 481 INKVAKYARLQSGFSGSRETLVSRHEIEVDPVKVHGTSYVFSDDGKISGDFAHRYA 540  
QY 541 SKCGIQTPTSAFQIRYGGYGVVDPDSMKLSLRKSMKSYESDNILKDLVLGWSKYQPC 600  
Db 541 SKCGIQTPTSAFQIRYGGYGVVDPDSMKLSLRKSMKSYESDNILKDLVLGWSKYQPC 600  
QY 601 YLNROLITLLSTGLVKDEVLEQKQEAVDQDLAILHDSLKQAQEALELMSPGENTINILKAM 660  
Db 601 YLNROLITLLSTGLVKDEVLEQKQEAVDQDLAILHDSLKQAQEALELMSPGENTINILKAM 660  
QY 661 LNCGYKPAEPLSMWLOTFRASKLLDLTRSRIPIPNGRTMWGCLDSRTLEYGQVFPVQ 720  
Db 661 LNCGYKPAEPLSMWLOTFRASKLLDLTRSRIPIPNGRTMWGCLDSRTLEYGQVFPVQ 720  
QY 721 FTGAGHGFEFDDLHPFNNSRSTNSFIKGNVYVAKNCPCLHPGDIRVLKAVNVRALHVMV 780  
Db 721 FTGAGHGFEFDDLHPFNNSRSTNSFIKGNVYVAKNCPCLHPGDIRVLKAVNVRALHVMV 780  
QY 781 DCWFVPQKGRPHNECSGSDLDGDIYFVCWDQDMIPPRVQPMYPPAPSIQLDHDVTI 840  
Db 781 DCWFVPQKGRPHNECSGSDLDGDIYFVCWDQDMIPPRVQPMYPPAPSIQLDHDVTI 840  
QY 841 EEVEEYFTNYIVNDSIGLIANAHHVVFADREPDMASDCKKLAELFSIADVFPKTPGPAE 900  
Db 841 EEVEEYFTNYIVNDSIGLIANAHHVVFADREPDMASDCKKLAELFSIADVFPKTPGPAE 900  
QY 901 IPSQLRPXEYDFMDPKDPTSYISERVIGKLFKRVKDKAPQASSIATPTRVARRSDAD 960  
Db 901 IPSQLRPXEYDFMDPKDPTSYISERVIGKLFKRVKDKAPQASSIATPTRVARRSDAD 960  
QY 961 MEVDGFEDYIDAFDYKTEYDNKLGNDYGIKTEAEILSGGIMKASKTFDRKDAEAI 1020  
Db 961 MEVDGFEDYIDAFDYKTEYDNKLGNDYGIKTEAEILSGGIMKASKTFDRKDAEAI 1020  
QY 1021 SVAVRALKEARAWFKRRNDIDMLPKASAWYHVTHYHTYWGCVNOGLKRAHFFSFPWCV 1080  
Db 1021 SVAVRALKEARAWFKRRNDIDMLPKASAWYHVTHYHTYWGCVNOGLKRAHFFSFPWCV 1080  
QY 1081 YDQLIQIKKDKARNRPVNLSSRLAQLSHRLVVK 1114  
Db 1081 YDQLIQIKKDKARNRPVNLSSRLAQLSHRLVVK 1114

RESULT 2  
T30828  
RNA-directed RNA polymerase - common tobacco  
C:Species: Nicotiana tabacum (common tobacco)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 20-Jun-2000  
C:Accession: T30828  
R:Schiebel, W.; Pelissier, T.; Riedel, L.; Thalmair, S.; Schiebel, R.; Kempe, D.; Lottsf  
submitted to the EMBL Data Library, October 1998  
A:Description: Isolation of an RNA-directed RNA polymerase-specific cDNA clone from Tom  
A:Reference number: Z20890  
A:Accession: T30828  
A:Molecule type: mRNA  
A:Residues: 1-1116 <SCH>  
A:Cross-references: EMBL:AJ011576; PIDN:CAA09697.1  
C:Genetics:  
A:Note: RdRP  
C:Superfamily: Arabidopsis probable RNA-directed RNA polymerase

Query Match 3.8%; Score 42; DB 2; Length 1116;  
Best Local Similarity 100.0%; Pred. No. 4.3e-36;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 979 EYDNKLGNDYGIKTEAEILSGGIMKASKTFDRKDAEAI 1020  
Db 981 EYDNKLGNDYGIKTEAEILSGGIMKASKTFDRKDAEAI 1022

RESULT 3  
T01920  
probable RNA-directed RNA polymerase (EC 2.7.7.48) - Arabidopsis thaliana  
N:Alternate names: protein T22B4.110  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 10-Dec-1999  
C:Accession: T01920; T08192  
R:Strong, C.; Graves, T.; Duckels, G.  
submitted to the EMBL Data Library, July 1998  
A:Description: The sequence of A. thaliana F2P3.  
A:Reference number: Z14455  
A:Accession: T01920  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1133 <STR>  
A:Cross-references: EMBL:AF080120; NID:g3600045; PID:g3600048  
A:Experimental source: cultivar Columbia  
R:Bevan, M.; Pohl, T.; Weizengger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Sch  
submitted to the Protein Sequence Database, May 1999  
A:Reference number: Z16098  
A:Accession: T08192  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1133 <BEV>  
A:Cross-references: EMBL:AL049876  
A:Experimental source: cultivar Columbia; BAC clone T22B4  
C:Genetics:  
A:Map position: 4  
A:Introns: 184/3; 820/2; 870/3  
A:Note: F2P3.11; T22B4.110  
C:Superfamily: Arabidopsis probable RNA-directed RNA polymerase  
C:Keywords: nucleotidyltransferase

Query Match 1.3%; Score 14; DB 2; Length 1133;  
Best Local Similarity 100.0%; Pred. No. 9.8e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 549 PSFQIRYGGYKGV 562  
Db 578 PSFQIRYGGYKGV 591

RESULT 4  
H69165  
hypothetical protein MTH500 - Methanobacterium thermoautotrophicum (strain Delta H)  
C:Species: Methanobacterium thermoautotrophicum  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 04-Mar-2000  
C:Accession: H69165  
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T  
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jilwani,  
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu  
A:Reference number: A69000; MUID:98037514; PMID:9371463  
A:Accession: H69165  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-583 <MTH>  
A:Cross-references: GB:AE000833; GB:AE000666; NID:g2621559; PIDN:AAB85006.1; PID:g262  
A:Experimental source: strain Delta H  
C:Genetics:  
A:Gene: MTH500  
C:Superfamily: Methanobacterium thermoautotrophicum hypothetical protein MTH500

Query Match 0.8%; Score 9; DB 2; Length 583;  
 Best Local Similarity 100.0%; Pred. No. 1.4; DB 2; Length 583;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 345 SPSITLDDG 353  
 |||||  
 DB 221 SPSITLDDG 229

RESULT 5  
 222763  
 hypothetical protein F56A8.2 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T22763  
 R:McMurray, A.  
 submitted to the EMBL Data Library, December 1996  
 A:Reference number: Z19612  
 A:Accession: T22763  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-230 <WTL>  
 A:Cross-references: EMBL:Z83230; PIDN:CAB05742.1; GSPDB:GN00021; CESP:F56A8.2  
 A:Experimental source: clone F56A8  
 C:Genetics:  
 A:Gene: CESP:F56A8.2  
 A:Map position: 3  
 A:Introns: 65/2; 97/3

Query Match 0.7%; Score 8; DB 2; Length 230;  
 Best Local Similarity 100.0%; Pred. No. 7.7; DB 2; Length 230;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 637 DSLKAQEA 644  
 |||||  
 DB 156 DSLKAQEA 163

RESULT 6  
 S60947  
 hypothetical protein YOR220w - yeast (Saccharomyces cerevisiae)  
 A:Alternate names: hypothetical protein O5050; hypothetical protein YOR50-10  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 15-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 19-Apr-2002  
 C:Accession: S60947; S67113; S71722  
 R:Galissou, F.; Dujon, B.  
 submitted to the EMBL Data Library, October 1995  
 A:Description: Sequence and analysis of a 33 kb fragment from the right arm of chromosome XV  
 A:Reference number: S60938  
 A:Accession: S60947  
 A:Molecule type: DNA  
 A:Residues: 1-265 <GAL>  
 A:Cross-references: EMBL:X92441; NID:g1050762; PID:g1050772  
 R:Boyer, J.; Fairhead, C.; Gailion, L.; Galissou, F.; Michaux, G.; Thierry, A.; Dujon, B.  
 submitted to the Protein Sequence Database, July 1996  
 A:Reference number: S67104  
 A:Accession: S67113  
 A:Molecule type: DNA  
 A:Residues: 1-265 <BOY>  
 A:Cross-references: EMBL:Z75128; NID:g1420509; PID:g1420510; MIPS:YOR220w  
 A:Experimental source: strain S288C  
 R:Galissou, F.; Dujon, B.  
 Yeast 12, 877-885, 1996  
 A:Title: Sequence and analysis of a 33 kb fragment from the right arm of chromosome XV  
 A:Reference number: S71722; MUID:96437977; PMID:8840505  
 A:Accession: S71722  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-265 <GAW>  
 A:Cross-references: EMBL:X92441; NID:g1050762; PIDN:CAA63183.1; PID:g1050772  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995  
 C:Genetics:  
 A:Cross-references: SGD:S0005746

A:Map position: 15R  
 C:Superfamily: Saccharomyces hypothetical protein YOR220w

Query Match 0.7%; Score 8; DB 2; Length 265;  
 Best Local Similarity 100.0%; Pred. No. 8.7; DB 2; Length 265;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 PPKSPSIT 349  
 |||||  
 DB 252 PPKSPSIT 259

RESULT 7  
 A95395  
 protein [imported] - Sinorhizobium meliloti (strain 1021) magaplasmid pSymA  
 C:Species: Sinorhizobium meliloti  
 C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
 C:Accession: A95395  
 R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; E.  
 ; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
 A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meli  
 A:Reference number: A95262; MUID:21396509; PMID:11481432  
 A:Accession: A95395  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-286 <KUR>  
 A:Cross-references: GB:AE006469; PIDN:AAK65723.1; PID:g14524217; GSPDB:GN00165  
 A:Experimental source: strain 1021, magaplasmid pSymA  
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl  
 pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.  
 L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau  
 hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,  
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A:Reference number: A96039; MUID:21368234; PMID:11474104  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: Sma1945  
 A:Genome: plasmid

Query Match 0.7%; Score 8; DB 2; Length 286;  
 Best Local Similarity 100.0%; Pred. No. 9.3; DB 2; Length 286;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 683 SKLIDLRT 690  
 |||||  
 DB 251 SKLIDLRT 258

RESULT 8  
 T38874  
 probable short chain dehydrogenase - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C:Accession: T38874  
 R:Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
 submitted to the EMBL Data Library, February 1996  
 A:Reference number: Z21732  
 A:Accession: T38874  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-341 <CON>  
 A:Cross-references: EMBL:Z69727; PIDN:CAA93565.1; GSPDB:GN000066; SPDB:SPAC4G9.15  
 A:Experimental source: strain 972h-; cosmid c4G9  
 C:Genetics:  
 A:Gene: SPDB:SPAC4G9.15  
 A:Map position: 1  
 A:Introns: 5/1; 48/3

Query Match 0.7%; Score 8; DB 2; Length 341;  
 Best Local Similarity 100.0%; Pred. No. 11;



Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1025 RALRKEAR 1032

Db 327 RALRKEAR 334

RESULT 9

H97272

histidinol-phosphate aminotransferase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001

C:Accession: H97272

R:Rolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: H97272

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-352 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK80971.1; PID:gl5026090; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC3031

C:Superfamily: histidinol phosphate aminotransferase

Query Match

Best Local Similarity 0.7%; Score 8; DB 2; Length 352;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 388 IDNFLRVS 395

Db 326 IDNFLRVS 333

RESULT 10

S51436

probable membrane protein YLR191w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein L9470.1

C:Species: Saccharomyces cerevisiae

C:Date: 23-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 06-Feb-1998

C:Accession: S51436

R:Wohldmann, P.

submitted to the EMBL Data Library, November 1994

A:Description: The sequence of S. cerevisiae cosmid 9470.

A:Reference number: S51414

A:Accession: S51436

A:Molecule type: DNA

A:Residues: 1-386 <WCH>

A:Cross-references: EMBL:U17246; NID:g577192; PID:g577193; MIPS:YLR191w

C:Genetics:

A:Gene: SGB:PEX13

A:Cross-references: SGD:S0004181; MIPS:YLR191w

A:Map position: 12R

C:Superfamily: SH3 homology

C:Keywords: transmembrane protein

F:264-280/Domain: transmembrane #status predicted <TM>

F:313-367/Domain: SH3 homology <SH3>

Query Match

Best Local Similarity 0.7%; Score 8; DB 2; Length 386;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 VFGFPYLL 14

Db 272 VFGFPYLL 279

RESULT 11

S69045

protein kinase MKK2 (EC 2.7.1.1) - yeast (Saccharomyces cerevisiae)

N:Alternate names: mitogen-activated protein kinase homolog; protein LPI6c; pr

C:Species: Saccharomyces cerevisiae

C:Date: 22-Aug-1996 #sequence\_revision 06-Sep-1996 #text\_change 24-Sep-1999

C:Accession: S69045; B48069; S30786

R:Hall, J.; DePaulo, T.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.

submitted to the EMBL Data Library, December 1995

A:Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.

A:Reference number: S69040

A:Accession: S69045

A:Molecule type: DNA

A:Residues: 1-506 <HAL>

A:Cross-references: EMBL:U43703; NID:gl244769; PIDN:AA68220.1; PID:gl244775; MIPS:YP

R:Rie, K.; Takase, M.; Lee, K.S.; Levin, D.E.; Araki, H.; Matsumoto, K.; Oshima, Y.

Mol. Cell. Biol. 13, 3076-3083, 1993

A:Title: MKK1 and MKK2, which encode Saccharomyces cerevisiae mitogen-activated prote

A:Reference number: A48069; MUID:93233668; PMID:8386320

A:Accession: B48069

A:Molecule type: DNA

A:Residues: 1-251, 'S', 253-506 <IRI>

A:Cross-references: EMBL:D13785; NID:g218442; PIDN:BA02933.1; PID:dl003439; PID:g218

A:Note: sequence extracted from NCBI backbone (NCBIP:129897)

C:Genetics:

A:Gene: SGD:MKK2; SSP33

A:Cross-references: SGD:S0006061; MIPS:YPL140C

A:Map position: 16L

C:Function:

A:Description: phosphotransferase; serine/threonine-specific protein kinase; involved

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom

C:Keywords: ATP; phosphotransferase; serine/threonine-specific protein kinase; signal

F:212-481/Domain: protein kinase homology <KIN>

F:220-228/Region: protein kinase ATP-binding motif

F:342/Active site: Asp #status predicted

Query Match

Best Local Similarity 0.7%; Score 8; DB 2; Length 506;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 923 ISERVIGK 930

Db 314 ISERVIGK 321

RESULT 12

G28771

hypothetical protein C2814 (photosynthetic gene cluster) - Rhodobacter capsulatus

C:Species: Rhodobacter capsulatus

C:Date: 18-Oct-1989 #sequence\_revision 18-Oct-1989 #text\_change 04-Sep-1998

C:Accession: G28771

R:Youvan, D.C.; Bylina, E.J.; Alberti, M.; Begusch, H.; Hearst, J.E.

Cell 37, 949-957, 1984

A:Title: Nucleotide and deduced polypeptide sequences of the photosynthetic reaction-

A:Reference number: A90850; MUID:84259352; PMID:6744416

A:Accession: G28771

A:Molecule type: DNA

A:Residues: 1-641 <YOUS>

A:Cross-references: GB:K01183

C:Superfamily: hypothetical protein C2814

Query Match

Best Local Similarity 0.7%; Score 8; DB 2; Length 641;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 465 GLTANDIR 472

Db 615 GLTANDIR 622

RESULT 13

T04848

protein kinase homolog F16G20.10 - Arabidopsis thaliana

N:Alternate names: protein F16G20.10; protein F21P8.200

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 17-Mar-2000

C:Accession: T04848; T05369  
 R:Bevan, M.; Massenet, O.; Clabault, G.; Quigley, F.; Mache, R.; Bancroft, I.; Mewes, H.  
 submitted to the Protein Sequence Database, August 1998  
 A:Reference number: Z15386  
 A:Accession: T04848  
 A:Molecule type: DNA  
 A:Residues: 1-830 <BEV>  
 A:Cross-references: EMBL:AL022347  
 A:Experimental source: cultivar Columbia; BAC clone F21P8  
 R:Bevan, M.; Watson, M.D.; Gibbons, T.; Bartley, J.; Bancroft, I.; Mewes, H.W.; Mayer, K.  
 submitted to the Protein Sequence Database, August 1998  
 A:Reference number: Z15413  
 A:Accession: T05369  
 A:Molecule type: DNA  
 A:Residues: 618-830 <BEW>  
 A:Cross-references: EMBL:AL031326  
 A:Experimental source: cultivar Columbia; BAC clone F16G20  
 C:Genetics:  
 A:Map position: 4  
 A:Introns: 484/1; 524/3; 595/1; 674/2; 725/3  
 A:Note: F21P8.200; F16G20.10  
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

Query Match 0.7%; Score 8; DB 2; Length 830;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 606 LITLSTL 613  
 |||||  
 Db 162 LITLSTL 169

RESULT 14  
 D70116  
 transcription elongation factor (greA) homolog - Lyme disease spirochete  
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
 C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 15-Oct-1999  
 C:Accession: D70116  
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White  
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,  
 Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
 Nature 390, 580-586, 1997  
 A:Authors: Smith, H.O.; Venter, J.C.  
 A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.  
 A:Reference number: A70100; MUID:98065943; PMID:9403685  
 A:Accession: D70116  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-901 <KLE>  
 A:Cross-references: GB:AE001124; GB:AE000783; NID:g2688003; PIDN:AAC66506.1; PID:g268800  
 A:Experimental source: strain B31  
 C:Keywords: transcription factor

Query Match 0.7%; Score 8; DB 2; Length 901;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 764 DIRVLKAV 771  
 |||||  
 Db 359 DIRVLKAV 366

RESULT 15  
 G70951  
 Probable ATP-dependent DNA helicase - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C:Accession: G70951  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete gen  
 A:Reference number: A70500; MUID:98295987; PMID:9634230  
 A:Accession: G70951  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1101 <COL>  
 A:Cross-references: GB:AL021646; GB:AL123456; NID:g3242278; PIDN:CAA16666.1; PID:el2  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: Rv3201c

Query Match 0.7%; Score 8; DB 2; Length 1101;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1021 SVAVRALR 1028  
 |||||  
 Db 374 SVAVRALR 381

RESULT 16  
 S34230  
 156K protein - Plantago asiatica mosaic virus  
 C:Species: Plantago asiatica mosaic virus  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 24-Nov-1999  
 C:Accession: S34230  
 R:Atabekov, J.G.  
 submitted to the EMBL Data Library, February 1993  
 A:Reference number: S34230  
 A:Accession: S34230  
 A:Status: preliminary  
 A:Molecule type: genomic RNA  
 A:Residues: 1-1395 <ATA>  
 A:Cross-references: EMBL:Z21647; NID:g311644; PIDN:CAA79761.1; PID:g311645  
 C:Superfamily: eggplant mosaic virus RNA-directed RNA polymerase

Query Match 0.7%; Score 8; DB 2; Length 1385;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 425 ERILSTLR 432  
 |||||  
 Db 597 ERILSTLR 604

RESULT 17  
 T27121  
 hypothetical protein Y53C10A.9 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
 C:Accession: T27121  
 R:White, S.  
 submitted to the EMBL Data Library, November 1998  
 A:Reference number: Z20314  
 A:Accession: T27121  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1564 <WIL>  
 A:Cross-references: EMBL:AL033536; PIDN:CAA22142.1; CESP:Y53C10A.9  
 A:Experimental source: clone Y53C10A  
 C:Genetics:  
 A:Gene: CESP:Y53C10A.9  
 A:Introns: 43/3; 92/2; 148/2; 226/3; 354/1; 712/3; 817/1; 875/1; 916/3; 984/3; 1059/2  
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog

Query Match 0.7%; Score 8; DB 2; Length 1564;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 606 LITLSTL 613  
 |||||  
 Db 254 LITLSTL 261

## RESULT 18

AB0717  
 hypothetical protein STY1874a [imported] - Salmonella enterica subsp. enterica serovar typhimurium  
 C:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A:Note: this species has also been called Salmonella typhi  
 C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
 C:Accession: AB0717  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Cronin, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; et al.  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar typhimurium  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: AB0717  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-29 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD02107.1; PID:g16502942; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY1874a

Query Match 0.6%; Score 7; DB 2; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 990 YGKTE 996

|||||||

Db 8 YGKTE 14

## RESULT 19

A57125  
 polyphenolic adhesive protein 3 - blue mussel  
 C:Species: Mytilus edulis (blue mussel)  
 C>Date: 03-Nov-1995 #sequence\_revision 03-Nov-1995 #text\_change 21-Jan-2000  
 C:Accession: A57125  
 R:Papov, V.V.; Diamond, T.V.; Biemann, K.; Waite, J.H.  
 J. Biol. Chem. 270, 20183-20192, 1995  
 A:Title: Hydroxyarginine-containing polyphenolic proteins in the adhesive plaques of the mussel Mytilus edulis  
 A:Reference number: A57125; MUID:95378278; PMID:7650037  
 A:Accession: A57125  
 A:Molecule type: protein  
 A:Residues: 1-48 <PAP>  
 F:3,4,8,14,19,22,25,29,32,47/Modified site: 3',4'-dihydroxyphenylalanine (Tyr) #status experimental  
 F:12,13,21,24,27,28,41,43,44/Modified site: 4-hydroxyarginine (Arg) #status experimental

Query Match 0.6%; Score 7; DB 2; Length 48;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 555 RVGGYKG 561

|||||||

Db 28 RVGGYKG 34

## RESULT 20

AB2371  
 hypothetical protein asr4522 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp.  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C:Accession: AB2371  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AB2371  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-66 <KUR>

QY 455 SVNMFAS 462

|||||||

Db 73 SVNMFAS 79

A:Cross-references: GB:BA000019; PIDN:BA076221.1; PID:g17133658; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: asr4522

Query Match 0.6%; Score 7; DB 2; Length 66;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 VQPPEGI 266

|||||||

Db 55 VQPPEGI 61

## RESULT 21

S68954  
 polyphenolic adhesive protein 3A precursor - Mediterranean mussel  
 N:Alternate names: foot protein 3A  
 C:Species: Mytilus galloprovincialis (Mediterranean mussel)  
 C>Date: 23-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 04-Feb-2000  
 C:Accession: S68954  
 R:Inoue, K.; Takeuchi, Y.; Miki, D.; Odo, S.; Harayama, S.; Waite, J.H.  
 Eur. J. Biochem. 239, 172-176, 1996  
 A:Title: Cloning, sequencing and sites of expression of genes for the hydroxyarginine  
 A:Reference number: S68954; MUID:96305382; PMID:8706704  
 A:Accession: S68954  
 A:Molecule type: mRNA  
 A:Residues: 1-70 <INO>

F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:25-70/Product: polyphenolic adhesive protein 3A #status predicted <MAT>  
 F:27,38,32,38,43,46,50,53,69,70/Modified site: 3',4'-dihydroxyphenylalanine (Tyr) #st  
 F:36,37,45,48,49,62,64,67/Modified site: 4-hydroxyarginine (Arg) #status predicted

Query Match 0.6%; Score 7; DB 2; Length 70;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 555 RVGGYKG 561

|||||||

Db 49 RVGGYKG 55

## RESULT 22

AD1030  
 probable membrane protein STY4565 [imported] - Salmonella enterica subsp. enterica se  
 C:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A:Note: this species has also been called Salmonella typhi  
 C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
 C:Accession: AD1030  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Cronin, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; et al.  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar typhimurium  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: AD1030  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-79 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD09341.1; PID:g16505341; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY4565

QY 455 SVNMFAS 462

|||||||

Db 73 SVNMFAS 79

Query Match 0.6%; Score 7; DB 2; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## RESULT 23

T46384  
 hypothetical protein DKF2p434p2119.1 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C>Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 02-Sep-2000  
 C:Accession: T46384  
 R:Ottewaeldeir, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, January 2000  
 A:Reference number: Z23031  
 A:Accession: T46384  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-89 <AAA>  
 A:Cross-references: EMBL:AL137543  
 A:Experimental source: adult testis; clone DKF2p434p2119  
 C:Genetics:  
 A:Note: DKF2p434p2119.1  
 C:Superfamily: tropomodulin

Query Match 0.6%; Score 7; DB 2; Length 89;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 652 ENTN1LK 658  
 |||||  
 Db 48 ENTN1LK 54

## RESULT 24

T07231  
 hypothetical protein 91 - Chlorella vulgaris chloroplast  
 C:Species: chloroplast Chlorella vulgaris  
 C>Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 21-Jul-2000  
 C:Accession: T07231  
 R:Wakasugi, T.; Nagai, T.; Kapoor, M.; Sugita, M.; Ito, M.; Ito, S.; Tsudzuki, J.; Nakas  
 Proc. Natl. Acad. Sci. U.S.A. 94, 5967-5972, 1997  
 A:Title: Complete nucleotide sequence of the chloroplast genome from the green alga Chlo  
 A:Reference number: Z15985; MUID:97303241; PMID:9159184  
 A:Accession: T07231  
 A:Status: preliminary; translated from GB/EMBL/DDDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-91 <WAK>  
 A:Cross-references: EMBL:AB001684; NID:g2224352; PIDN:BAAS7878.1; PID:g2224394  
 C:Genetics:  
 A:Genome: chloroplast  
 C:Keywords: chloroplast

Query Match 0.6%; Score 7; DB 2; Length 91;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 ASRPGLT 467  
 |||||  
 Db 6 ASRPGLT 12

## RESULT 25

SI4355  
 glutathione transferase (EC 2.5.1.18) class alpha chain 10 - rat (fragments)  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 21-Nov-1993 #sequence\_revision 08-Nov-1996 #text\_change 08-Nov-1996  
 C:Accession: SI4355  
 R:Meyer, D.J.; Gilmore, K.S.; Coles, B.; Dalton, K.; Hulbert, P.B.; Ketterer, B.  
 Biochem. J. 274, 619, 1991  
 A:Title: Structural distinction of rat GSH transferase subunit 10.  
 A:Reference number: SI4355; MUID:91174781; PMID:2006926  
 A:Accession: SI4355  
 A:Molecule type: protein  
 A:Residues: 1-29;30-51;52-60;61-66;67-94;95-112 <MEY>  
 C:Superfamily: glutathione transferase  
 C:Keywords: dimer; transferase

Query Match 0.6%; Score 7; DB 2; Length 112;

Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095  
 |||||  
 Db 79 KDKARNR 85

## RESULT 26

AG2329  
 50S ribosomal protein L17 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp.  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C:Accession: AG2329  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigi  
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AG2329  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-116 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BAE75889.1; PID:g17133325; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: rpl17  
 C:Superfamily: Escherichia coli ribosomal protein L17

Query Match 0.6%; Score 7; DB 2; Length 116;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 DKIIITLA 65  
 |||||  
 Db 49 DKIIITLA 55

## RESULT 27

D97840  
 hypothetical protein RC1124 [imported] - Rickettsia conorii (strain Malish 7)  
 C:Species: Rickettsia conorii  
 C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 30-Sep-2001  
 C:Accession: D97840  
 R:Ogata, H.; Audic, S.; Repesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.;  
 Science 293, 2093-2098, 2001  
 A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
 A:Reference number: A97700; MUID:21442074; PMID:11557893  
 A:Accession: D97840  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-121 <KUR>  
 A:Cross-references: GB:AE006914; PIDN:AAL03662.1; PID:g15620249; GSPDB:GN00173  
 C:Genetics:  
 A:Gene: RC1124

Query Match 0.6%; Score 7; DB 2; Length 121;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KTIQVFG 9  
 |||||  
 Db 33 KTIQVFG 39

## RESULT 28

AC2172  
 hypothetical protein all2930 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp.  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C:Accession: AC2172

R.Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shampo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena dactyloides* strain 782-874-2

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AC2172

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-122 <KUR>

A:Cross-references: GB:BA000013; PIDN:BA074629.1; PID:gl17132024; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all2930

Query Match 0.6%; Score 7; DB 2; Length 122;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 VLRNYS 386  
|||||||

Db 107 VLRNYS 113

RESULT 29

AG1754

C:Species: *Listeria innocua*

C:Title: bacteriophage protein homolog lin2580 [imported] - *Listeria innocua* (strain Clip11262)

C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C:Accession: AG1734

R:Glaser, P.; Fraungel, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, R.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H. D.; Jones, L.M.; Karst, U. Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunz, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlund, A.; Title: Comparative genomics of *Listeria species*

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AG1734

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-124 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC97807.1; PID:gl16415102; GSPDB:GN00178

A:Experimental source: strain Clip11262

C:Genetics:

A:Gene: lin2580

Query Match 0.6%; Score 7; DB 2; Length 124;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 976 YKTEYDN 982  
|||||||

Db 32 YKTEYDN 38

RESULT 30

H71504

C:Species: *Chlamydia trachomatis*

C:Title: Ribosomal protein L17 - *Chlamydia trachomatis*

C>Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 13-Aug-1999

C:Accession: H71504; I40747

R:Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis* serotype D

A:Reference number: A71570; MUID:99000809; PMID:97841136

A:Accession: H71504

A:Molecule type: DNA

A:Residues: 1-141 <ARN>

A:Cross-references: GB:AB001323; GB:AB001273; NID:g3328931; PIDN:AAC68107.1; PID:g332894

A:Experimental source: serotype D, strain UW-3/Cx

R:Gu, L.; Wernan, W.M.; Remacha, M.; Meuser, R.; Coffin, J.; Kaul, R. J. Bacteriol. 177, 2594-2601, 1995

A:Title: *Chlamydia trachomatis* RNA polymerase alpha subunit: sequence and structural analysis

A:Reference number: I40743; MUID:95247702; PMID:7730299

A:Accession: I40747

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 20-115,'R',117-141 <GUL>

A:Cross-references: GB:L33834; NID:g620026; PIDN:AAA74990.1; PID:g620030

C:Genetics:

A:Gene: rli7

C:Superfamily: Escherichia coli ribosomal protein L17

C:Keywords: protein biosynthesis; ribosome

Query Match 0.6%; Score 7; DB 2; Length 141;  
Best Local Similarity 100.0%; Pred. No. 61;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 926 RVIGKLF 932  
|||||||

Db 98 RVIGKLF 104

RESULT 31

AB1663

C:Species: *Chlamydia muridarum* (strain Nigg)

C:Title: ribosomal protein L17 TC0793 [imported] - *Chlamydia muridarum* (strain Nigg)

C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 02-Sep-2000

C:Accession: AB1663

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hicke, C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39

A:Reference number: AB1500; MUID:20150255; PMID:10684935

A:Accession: AB1663

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-142 <RET>

A:Cross-references: GB:AE002347; GB:AE002160; NID:g7190815; PIDN:AAF39596.1; PID:g719

A:Experimental source: strain Nigg (MoPn)

C:Genetics:

A:Gene: TC0793

C:Superfamily: Escherichia coli ribosomal protein L17

Query Match 0.6%; Score 7; DB 2; Length 142;  
Best Local Similarity 100.0%; Pred. No. 61;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 926 RVIGKLF 932  
|||||||

Db 98 RVIGKLF 104

RESULT 32

H82567

C:Species: *Xylella fastidiosa*

C:Title: outer membrane protein XF2345 [imported] - *Xylella fastidiosa* (strain 9A5c)

C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C:Accession: H82567

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A:Reference number: AB2515; MUID:20365717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: H82567

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-145 <SIM>

A:Cross-references: GB:AE004045; GB:AE003849; NID:g9107517; PIDN:AAF95144.1; GSPDB:GN

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R. Briones, M.R.S.; Bueno, M.R.P.; Camargo, L.E.A.; Carraro, D.M.; Carrier as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S. submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fi J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; La

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.; Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A.; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tshako, M.H.; Valiada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A;Reference number: A59328  
C;Contents: annotation  
C;Genetics:  
A;Gene: XF2345

Query Match 0.6%; Score 7; DB 2; Length 145;  
Best Local Similarity 100.0%; Pred. No. 62;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 625 KEAVDQL 631  
Db 49 KEAVDQL 55

## RESULT 33

B89929  
hypothetical protein SA1329 [imported] - Staphylococcus aureus (strain N315)  
C;Species: Staphylococcus aureus  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C;Accession: B89929

R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kato, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001

A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A;Reference number: A89758; MUID:21311952; PMID:11418146

A;Accession: B89929  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-149 <KUR>

A;Cross-references: GB:BA000018; PID:gl3701297; PIDN:BA842591.1; GSPDB:GN00149  
A;Experimental source: strain N315

C;Genetics:

A;Gene: SA1329  
C;Superfamily: ferric uptake regulator

Query Match 0.6%; Score 7; DB 2; Length 149;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 934 KVDDKAP 940  
Db 46 KVDDKAP 52

## RESULT 34

A60943  
Ig heavy chain precursor V region (clone HN.14) - human

C;Species: Homo sapiens (man)  
C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 17-Mar-1999  
C;Accession: A60943; A48165  
R;Desai, R.; Spatz, L.; Matsuda, T.; Ilyas, A.A.; Berman, J.E.; Alt, F.W.; Kabat, E.A.; J. Neuroimmunol. 30, 245, 1990

A;Title: Molecular cloning of a human immunoglobulin heavy chain variable (V-H) region  
A;Reference number: A60943; MUID:91036050; PMID:1699376

A;Accession: A60943  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-151 <DES>

J. Desai, R.; Spatz, L.; Matsuda, T.; Ilyas, A.A.; Berman, J.E.; Alt, F.W.; Kabat, E.A.; J. Neuroimmunol. 26, 35-41, 1990  
A;Title: Molecular cloning of a human immunoglobulin heavy chain variable (V-H) region  
A;Reference number: A48165; MUID:90094677; PMID:1688442

A;Accession: A48165  
A;Molecule type: mRNA  
A;Residues: 1-36, 'M', '38-62, 'AR', '67-151 <DE2>

A;Note: this sequence has been corrected in reference A60943

C;Superfamily: immunoglobulin V region; immunoglobulin homology  
C;Keywords: heterotetramer; immunoglobulin  
F:34-117/Domain: immunoglobulin homology <IMM>

Query Match 0.6%; Score 7; DB 2; Length 151;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 ASTGSGV 419  
Db 122 ASTGSGV 128

## RESULT 35

S75019  
glutathione peroxidase (EC 1.11.1.9) - Synechocystis sp. (strain PCC 6803)  
N;Alternate names: hypothetical protein slr1992  
C;Species: Synechocystis sp.  
A;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis S.

A;Reference number: S74322; MUID:97061201; PMID:8905231  
A;Accession: S75019  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-154 <KAN>

A;Cross-references: EMBL:D90910; GB:AB001339; NID:gl652956; PIDN:BAAL7881.1; PID:gl165  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C;Superfamily: glutathione peroxidase  
C;Keywords: oxidoreductase; selenocysteine  
F:34/Modified site: selenocysteine #status predicted

Query Match 0.6%; Score 7; DB 2; Length 154;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKGGL 545  
Db 30 VASKGGL 36

## RESULT 36

H64930  
Hypothetical protein b1720 - Escherichia coli (strain K-12)

C;Species: Escherichia coli  
C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C;Accession: H64930  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; A.; Ross, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.  
A;Reference number: A54720; MUID:97426617; PMID:9278503

A;Accession: H64930  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA  
A;Residues: 1-157 <BLAT>

A;Cross-references: GB:AE000267; GB:U00096; NID:gl788011; PIDN:AAAC74790.1; PID:gl7880  
A;Experimental source: strain K-12, substrain MG1655  
C;Superfamily: Escherichia coli hypothetical protein b1720

Query Match 0.6%; Score 7; DB 2; Length 157;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 604 ROLITLL 610  
Db 37 ROLITLL 43

RESULT 37  
F83541  
Probable glutathione peroxidase PA0838 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: F83541  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B.  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: F83541  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-160 <STO>  
A:Cross-references: GB:AE004518; GB:AE004091; NID:99946725; PIDN:AAG04227.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0838  
C:Superfamily: glutathione peroxidase

Query Match 0.6%; Score 7; DB 2; Length 160;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKCGL 545  
|||||  
DB 32 VASKCGL 38

RESULT 38  
T14262  
glutathione peroxidase (EC 1.11.1.9) - common sunflower  
C:Species: Helianthus annuus (common sunflower)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Jun-2000  
C:Accession: T14262  
R:Roedel-Drevet, P.; Gagne, G.; Tourvielle de Labrouhe, D.; Dufaure, J.P.; Nicolas, P.;  
Physiol. Plantarum 103, 385-394, 1998  
A:Title: Molecular characterization, organ distribution and stress-mediated induction of  
A:Reference number: 217562  
A:Accession: T14262  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-167 <ROE>  
A:Cross-references: EMBL:Y14429  
C:Genetics:  
A:Gene: GPxha-1  
C:Function:  
A:Description: catalyzes reduction of hydroperoxides by glutathione, thus protecting bio  
C:Superfamily: glutathione peroxidase  
C:Keywords: oxidoreductase

Query Match 0.6%; Score 7; DB 2; Length 167;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKCGL 545  
|||||  
DB 37 VASKCGL 43

RESULT 39  
D84722  
Probable glutathione peroxidase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: D84722  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Woffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: D84722  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-169 <STO>  
A:Cross-references: GB:AE002093; NID:94582452; PIDN:AD24836.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g31570  
A:Map position: 2  
C:Superfamily: glutathione peroxidase

Query Match 0.6%; Score 7; DB 2; Length 169;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKCGL 545  
|||||  
DB 37 VASKCGL 43

RESULT 40  
ORECBE  
Vitamin B12 transport periplasmic protein btuE - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 30-Jun-1988 #sequence\_revision 31-Oct-1997 #text\_change 01-Mar-2002  
C:Accession: F64929; B24498  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: F64929  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-183 <BLAT>  
A:Cross-references: GB:AE000266; GB:U00096; NID:q1787997; PIDN:AAC74780.1; PID:q17880  
A:Experimental source: strain K-12, substrain MG1655  
R:Friedrich, M.J.; Deveau, L.C.; Kadner, R.J.  
J. Bacteriol. 167, 928-934, 1986  
A:Title: Nucleotide sequence of the btuCED genes involved in vitamin B12 transport in  
A:Reference number: A24498; MUID:86304184; PMID:3528129  
A:Accession: B24498  
A:Molecule type: DNA  
A:Residues: 1-183 <FR>  
A:Cross-references: GB:M14031; NID:q145441; PIDN:AAA23527.1; PID:q145444  
C:Comment: This protein, thought to be located in the periplasm, is not essential for  
C:Comment: This sequence is homologous with that of bovine glutathione peroxidase (41  
C:Genetics:  
A:Gene: btuE  
A:Map position: 37 min  
C:Superfamily: glutathione peroxidase  
C:Keywords: vitamin B12 transport  
F:37/Active site: Cys #status predicted

Query Match 0.6%; Score 7; DB 1; Length 183;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKCGL 545  
|||||  
DB 33 VASKCGL 39

RESULT 41  
A98931  
vitamin B12 transport protein ECs2417 [imported] - Escherichia coli (strain O157:H7,  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: A98931  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic islands  
A:Reference number: A98629; MUID:21156231; PMID:11256796  
A:Accession: A98931  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-183 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA035840.1; PID:g13361884; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RMD 0509952  
C:Genetics:  
A:Gene: ECs2417  
C:Superfamily: glutathione peroxidase

Query Match 0.6%; Score 7; DB 2; Length 183;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKCGL 545  
|||||  
Db 33 VASKCGL 39

RESULT 42  
E85779  
vitamin B12 transport [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: E85779  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: E85779  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-183 <STO>  
A:Cross-references: GB:AE005174; NID:g12515721; PIDN:AAG56697.1; GSPDB:GN00145; UNCP:227  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: btuE  
C:Superfamily: glutathione peroxidase

Query Match 0.6%; Score 7; DB 2; Length 183;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKCGL 545  
|||||  
Db 33 VASKCGL 39

RESULT 43  
S04743  
TPA-induced protein 11 - mouse  
N:Alternate names: tetradecanoyl phorbol acetate-induced protein 11  
C:Species: Mus musculus (house mouse)  
C:Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 19-Apr-2002  
C:Accession: S04743  
R:Varum, B.C.; Lim, R.W.; Sukhatme, V.P.; Herschman, H.R.  
Oncogene 4, 119-120, 1989  
A:Title: Nucleotide sequence of a cDNA encoding TIS11, a message induced in Swiss 3T3 ce  
A:Reference number: S04743; MUID:89128189; PMID:2915901  
A:Accession: S04743  
A:Molecule type: mRNA  
A:Residues: 1-183 <VAR>  
A:Cross-references: EMBL:X14678; NID:g54803; PIDN:CAA32807.1; PID:g54804  
C:Genetics:  
A:Gene: TIS11  
C:Superfamily: Caenorhabditis elegans hypothetical protein Y116A8C.20  
C:Keywords: DNA binding; phosphoprotein; zinc finger

Query Match 0.6%; Score 7; DB 2; Length 183;  
Best Local Similarity 100.0%; Pred. No. 76;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 ILFKISS 278  
|||||  
Db 9 ILFKISS 15

RESULT 44  
TI5923  
hypothetical protein EED8.12 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: TI5923  
R:Chissoe, S.  
submitted to the EMBL Data Library, July 1995  
A:Description: The sequence of C. elegans cosmid EED8.  
A:Reference number: Z18428  
A:Accession: TI5923  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-197 <CHI>  
A:Cross-references: EMBL:U23484; NID:g733597; PID:g733609; PIDN:AAC46772.1; CESP:EEDC  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Gene: CESP:EED8.12  
A:Introns: 103/2

Query Match 0.6%; Score 7; DB 2; Length 197;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 839 TIEVEE 845  
|||||  
Db 73 TIEVEE 79

RESULT 45  
A84865  
probable glutathione peroxidase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: A84865  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: A84865  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-206 <STO>  
A:Cross-references: GB:AE002093; NID:g2289006; PIDN:AAB64335.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g43350  
A:Map position: 2  
C:Superfamily: glutathione peroxidase

Query Match 0.6%; Score 7; DB 2; Length 206;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKCGL 545  
|||||  
Db 76 VASKCGL 82

RESULT 46  
E75310  
hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 28-Jul-2000  
C:Accession: E75310



R.White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896; PMID:10567266  
 A:Accession: E75310  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-206 <WHI>  
 A:Cross-references: GB:AE002048; GB:AE000513; NID:g6459929; PIDN:AAF11690.1; PID:g645993  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR2136  
 A:Map position: 1  
 C:Superfamily: Deinococcus radiodurans hypothetical protein DR2136

Query Match 0.6%; Score 7; DB 2; Length 206;  
 Best Local Similarity 100.0%; Pred. No. 85;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 959 ADMEVDG 965  
 |||||  
 Db 140 ADMEVDG 146

RESULT 47  
 F83815  
 hypothetical protein BH1326 [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans  
 C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C:Accession: F83815  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A:Reference number: A83650; MUID:20512582; PMID:11058132  
 A:Accession: F83815  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-207 <STO>  
 A:Cross-references: GB:AF001511; GB:BA000004; NID:g10173727; PIDN:BA05045.1; GSPDB:GNOC  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH1326  
 C:Superfamily: Bacillus subtilis conserved hypothetical protein yqeJ

Query Match 0.6%; Score 7; DB 2; Length 207;  
 Best Local Similarity 100.0%; Pred. No. 85;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 GPEVNV 377  
 |||||  
 Db 152 GPEVNV 158

RESULT 48  
 JC2368  
 ribosomal protein L13, cytosolic [validated] - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000  
 C:Accession: JC2368; PC2232; JC2235; PC2157  
 R:Olvera, J.; Wool, I.G.  
 Biochem. Biophys. Res. Commun. 201, 102-107, 1994  
 A:Title: The primary structure of rat ribosomal protein L13.  
 A:Reference number: JC2235; MUID:94256964; PMID:8198561  
 A:Accession: JC2368  
 A:Molecule type: mRNA  
 A:Residues: 1-211 <OLV1>  
 A:Cross-references: EMBL:X78327; NID:g510551; PIDN:CAA55130.1; PID:g510552  
 A:Accession: PC2232  
 A:Molecule type: protein  
 A:Residues: 2-14;26-60;100-180 <OLV2>  
 A:Experimental source: clone pLI3-2,3

A:Note: the protein is designated as ribosomal protein L13  
 C:Superfamily: rat ribosomal protein L13  
 C:Keywords: protein biosynthesis; ribosome  
 F;2-211/Product: ribosomal protein L13 #status predicted <WAT>

Query Match 0.6%; Score 7; DB 1; Length 211;  
 Best Local Similarity 100.0%; Pred. No. 86;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 VDPRRRN 304  
 |||||  
 Db 98 VDPRRRN 104

RESULT 49  
 S23753  
 ribosomal protein L13, cytosolic - human  
 N:Alternate names: BBCL protein  
 C:Species: Homo sapiens (man)  
 C:Date: 27-May-1994 #sequence\_revision 10-Nov-1995 #text\_change 13-Aug-1999  
 C:Accession: S23753  
 R:Adams, S.M.; Helps, N.R.; Sharp, M.G.F.; Brammar, W.J.; Walker, R.A.; Varley, J.M.  
 Hum. Mol. Genet. 1, 91-96, 1992  
 A:Title: Isolation and characterization of a novel gene with differential expression  
 A:Reference number: S23753; MUID:93244791; PMID:1301162  
 A:Accession: S23753  
 A:Molecule type: mRNA  
 A:Residues: 1-211 <ADA>  
 A:Cross-references: EMBL:X64707; NID:g29382; PIDN:CAA45963.1; PID:g29383  
 C:Superfamily: rat ribosomal protein L13  
 C:Keywords: cytosol; protein biosynthesis; ribosome

Query Match 0.6%; Score 7; DB 2; Length 211;  
 Best Local Similarity 100.0%; Pred. No. 86;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 VDPRRRN 304  
 |||||  
 Db 98 VDPRRRN 104

RESULT 50  
 T22984  
 hypothetical protein F59B8.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T22984  
 R:Matthews, P.  
 submitted to the EMBL Data Library, January 1996  
 A:Reference number: Z19647  
 A:Accession: T22984  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-213 <WIL>  
 A:Cross-references: EMBL:Z68343; PIDN:CAA92779.1; GSPDB:GN00022; CESP:F59B8.1  
 A:Experimental source: clone F59B8  
 C:Genetics:  
 A:Gene: CBSP:F59B8.1  
 A:Map position: 4  
 A:Introns: 31/2; 65/3; 94/2; 138/1; 170/3

Query Match 0.6%; Score 7; DB 2; Length 213;  
 Best Local Similarity 100.0%; Pred. No. 87;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 LHDSLKA 641  
 |||||  
 Db 132 LHDSLKA 138

Search completed: November 6, 2002, 03:39:12  
 Job time : 91 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:26:03 ; Search time 87 Seconds  
(without alignments)  
531.088 Million cell updates/sec

Title: US-09-782-874-2  
Perfect score: 1114  
Sequence: 1 MGKTIQVGFPPYLLSAEYWK.....RPVNLSSLRQAQLSHRLVLK 1114

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	9	0.8	2258	1 FA5_PIG	Q9g1p1 sus scrofa
2	8	0.7	341	1 YDLF_SCHPO	Q10245 schizosacch
3	8	0.7	352	1 HIS8_CLOAB	Q97es6 clostridium
4	8	0.7	386	1 PXND_YEAST	P80667 saccharomyc
5	8	0.7	506	1 MKX2_YEAST	P32491 saccharomyc
6	8	0.7	641	1 DXS_RHOCA	P26242 rhodobacter
7	8	0.7	901	1 GREX_BORBU	O31157 borrelia bu
8	8	0.7	1283	1 PEX1_HUMAN	O43933 homo sapien
9	8	0.7	1385	1 RRPO_PIAWV	Q07518 plantago as
10	7	0.6	90	1 REV_HV128	P05869 human immun
11	7	0.6	141	1 RL17_CHLTR	P47760 chlamydia t
12	7	0.6	142	1 RL17_CHLMU	Q9pjn5 chlamydia m
13	7	0.6	148	1 CYN5_SULTO	Q972w5 sulfolobus
14	7	0.6	167	1 GSHZ_HELAN	O33970 helianthus
15	7	0.6	169	1 GSHX_ARATH	O04922 arabidopsis
16	7	0.6	183	1 BTUE_ECOLI	P06610 escherichia
17	7	0.6	197	1 YQOC_CAEEL	Q93301 caenorhabdi
18	7	0.6	207	1 NADD_BACHD	Q9kd91 bacillus ha
19	7	0.6	210	1 RL13_CHICK	P41125 gallus gall
20	7	0.6	210	1 RL13_CRIGR	Q92133 cricetus
21	7	0.6	210	1 RL13_HUMAN	P26373 homo sapien
22	7	0.6	210	1 RL13_ICTPU	Q90yv5 ictalurus p
23	7	0.6	210	1 RL13_MOUSE	P47963 mus musculu
24	7	0.6	210	1 RL13_RAT	P41123 rattus norv
25	7	0.6	214	1 YDHI_HSVS7	P25049 herpesvirus
26	7	0.6	219	1 RL13_SPOPR	Q962u1 spodopvira
27	7	0.6	220	1 GTC1_RAT	P04904 rattus norv
28	7	0.6	220	1 GTC2_RAT	P46418 rattus norv
29	7	0.6	229	1 KPIK_THELI	Q56301 thermococcu
30	7	0.6	229	1 EUTQ_SALTY	Q92iv5 salmonella
31	7	0.6	231	1 RADC_BACSU	Q02170 bacillus su
32	7	0.6	239	1 I431_ENTHI	P42648 entamoeba h
33	7	0.6	274	1 CCS_MOUSE	Q9wu84 mus musculu

34	7	0.6	276	1 UPK_BACSU	P94507 bacillus su
35	7	0.6	291	1 Y620_METJA	Q58037 methanococc
36	7	0.6	303	1 UL07_HSVBE	P28945 equine herp
37	7	0.6	307	1 GLSA_STRCO	P57755 streptomyce
38	7	0.6	311	1 LIP_PSPSP	P26877 pseudomonas
39	7	0.6	314	1 ISPH_BACHD	Q9kd37 bacillus ha
40	7	0.6	320	1 VG2_SPV4	P11334 spiroplasma
41	7	0.6	329	1 Y526_CHLPN	Q92826 chlamydia p
42	7	0.6	332	1 PDXA_FUSNN	Q8rgro fusobacteri
43	7	0.6	333	1 RPOA_PSEAE	O52760 pseudomonas
44	7	0.6	335	1 YO94_CAEEL	P41844 caenorhabdi
45	7	0.6	342	1 RUVE_RICCN	Q92187 rickettsia
46	7	0.6	342	1 RUVE_RICPR	Q92ae5 rickettsia
47	7	0.6	345	1 NQ08_PARDE	P29920 paracoccus
48	7	0.6	358	1 YJ9L_YEAST	P47175 saccharomyc
49	7	0.6	362	1 BIOB_SYNY3	P73538 synchocyst
50	7	0.6	371	1 YPC4_CAEEL	Q11181 caenorhabdi

#### ALIGNMENTS

RESULT 1  
FA5\_PIG ID FA5\_PIG STANDARD; PRT; 2258 AA.  
AC Q9G1P1;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Coagulation factor V precursor (Activated protein C cofactor).  
GN F5.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID:9623;  
RN [1]  
RP SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING OF F5/8 TYPE A AND C  
DOAINS.

TISSUE=Liver;  
MEDLINE=21121490; PubMed=11229814;  
RA Grimm D.R., Colter M.B., Braunschweig M., Alexander L.J., Neame P.J.,  
RA Kim H.K.W.;  
RT "Porcine factor V: cDNA cloning, gene mapping, three-dimensional  
protein modeling of membrane binding sites and comparative anatomy of  
domains.";  
RL Cell. Mol. Life Sci. 58:148-159(2001).  
CC -!- FUNCTION: Coagulation factor V is a cofactor that participates  
with factor Xa to activate prothrombin to thrombin.  
CC -!- SUBUNIT: Factor Va is composed of a heavy chain and a light  
chain, noncovalently bound. The interaction between the two chains  
is calcium-dependent.  
CC -!- DOMAIN: Domain B contains 41 x 9 AA tandem repeats. Domains C1  
and C2 may be involved in membrane binding.  
CC -!- PM: Thrombin activates factor V proteolytically to the active  
cofactor, factor Va (formation of a heavy chain at the N-  
terminus and a light chain at the C-terminus).  
CC -!- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF  
2 PLASTOCYANIN-LIKE REPEATS.  
CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF191308; AAC28381.1; -  
DR HSSP; P12259; ICZT.  
DR InterPro; IPR001117; Cu-oxidase.  
DR InterPro; IPR000421; FA58\_C.  
DR



RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Gobie A., Hamlin N., Harris D., Hidalgo J., Hodges G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckart G., Aert R., Robben J., Grymonprez B.,  
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovsky G.V., Ussery D., Barrell B.G., Nurse P.,  
RT "The genome sequence of Schizosaccharomyces pombe";  
RL Nature 415:871-880(2002).  
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
CC (SDR) FAMILY. STRONG, TO YEAST YBR159W.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z69727; CAA93565.1; -;  
DR InterPro: IPR002198; ADH\_short.  
DR Pfam: PF00166; adh\_short; 1.  
DR PROSITE: PS00061; ADH\_SHORT; 1.  
KW Hypothetical protein; Oxidoreductase; Transmembrane.  
FT TRANSMEM 11 31 POTENTIAL.  
FT ACT\_SITE 216 216 BY SIMILARITY.  
SQ SEQUENCE 341 AA; 37307 MW; F49CAAB79486F71D CRC64;  
  
Query Match 0.7%; Score 8; DB 1; Length 341;  
Best Local Similarity 100.0%; Pred. No. 4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1025 RALRKEAR 1032  
DB 327 RALRKEAR 334  
|||||||  
  
RESULT 3  
HIS8\_CLOB STANDARD; PRT; 352 AA.  
AC Q97B56;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Histidinol-phosphate aminotransferase (EC 2.6.1.9) (Imidazole acetol-  
DE phosphate transaminase).  
GN HSC OR CAC3031.  
OS Clostridium acetobutylicum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1488;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
RX MEDLINE=21359325; PubMed=11466286;  
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
RA Bennett G.N., Koonin E.V., Smith D.R.;  
RT "Genome sequence and comparative analysis of the solvent-producing  
RT bacterium Clostridium acetobutylicum";  
RL J. Bacteriol. 183:4823-4838(2001).  
CC -!- CATALYTIC ACTIVITY: L-histidinol-phosphate + 2-oxoglutarate = 3-  
CC (imidasol-4-yl)-2-oxopropyl phosphate + L-glutamate.  
CC -!- COFACTOR: Pyridoxal phosphate (By similarity).  
CC -!- PATHWAY: Histidine biosynthesis; seventh step.  
CC -!- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
CC AMINOTRANSFERASES. HISTIDINOL-PHOSPHATE AMINOTRANSFERASES  
CC SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AE007800; AAK80971.1; -;  
DR InterPro: IPR004839; Aminotransf1/2.  
DR InterPro: IPR001917; NHtransf\_2.  
DR Pfam: PF00155; aminotran\_1.2; 1.  
DR TIGRFAMS: TIGR01141; hisc; 1.  
DR PROSITE: PS00599; AA-TRANSFER\_CLASS\_2; 1.  
KW Histidine biosynthesis; Transferase; Aminotransferase;  
KW Pyridoxal phosphate; Complete proteome.  
FT BINDING 210 210 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
SQ SEQUENCE 352 AA; 40220 MW; 3375021EBEC6BC2E CRC64;  
  
Query Match 0.7%; Score 8; DB 1; Length 352;  
Best Local Similarity 100.0%; Pred. No. 4.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 388 IDNFLRVS 395  
DB 326 IDNFLRVS 333  
|||||||  
  
RESULT 4  
PEXD\_YEAST STANDARD; PRT; 386 AA.  
AC P80667;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Peroxisomal membrane protein PAS20 (Peroxin-13).  
GN PEX13 OR PAS20 OR YLR191W OR L9470.1.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97011156; PubMed=8858166;  
RA Elgersma Y., Kwast L., Klein A., Voorn-Brouwer T., van den Berg M.,  
RA Tabak H.F., Distel B.;  
RT "The SH3 domain of the Saccharomyces cerevisiae peroxisomal membrane  
RT protein Pex13p functions as a docking site for Pex5p, a mobile  
RT receptor for the import PTS1-containing proteins";  
RL J. Cell Biol. 135:97-109(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / AB972;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,  
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,

```

RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
RA Tach A., Travaskis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
RA Wilson R., Waterston R.;
RA Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC STRAIN=S288C;
RX MEDLINE=97011157; PubMed=8858167;
RA Erdmann R., Blobel G.;
RA "Identification of Pex13p a peroxisomal membrane receptor for the
RT Pex1 recognition factor.";
RL J. Cell Biol. 135:111-121(1996).
CC -1- FUNCTION: COMPONENT OF THE PEROXISOMAL TRANSLOCATION MACHINERY
CC WITH PEX14 AND PEX17. INTERACTS WITH THE PEX13 RECEPTOR
CC (PAS10/PEX5).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Peroxisomal.
CC . -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; S82971; AAB46885.1; -
CC EMBL; U37420; AAA79308.1; -
CC EMBL; U17246; AAB67453.1; -
CC EMBL; U14913; AAB67448.1; -
CC HSP; Q06187; IAWX.
CC SGD; S0004181; PEX13.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00018; SH3; 1.
CC PRINTS; PR00452; SH3DOMAIN.
CC ProDom; PD000066; SH3; 1.
CC SMART; SM00326; SH3; 1.
CC PROSITE; PS50002; SH3; 1.
KW Peroxisome; Transport; Protein transport; Transmembrane; SH3 domain.
FT DOMAIN 1 263 LUMENAL (POTENTIAL).
FT TRANSMEM 264 280 POTENTIAL.
FT DOMAIN 281 386 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 306 372 SH3.
SQ SEQUENCE 386 AA; 42706 MW; DBEA9A2372185860 CRC64;

Query Match 0.7%; Score 8; DB 1; Length 386;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 VFQPPYLL 14
| | | | | | |
Db 272 VFQPPYLL 279

RESULT 5
MKK2 YEAST STANDARD; PRT; 506 AA.
AC P32491.
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE MAP kinase kinase MKK2/SSP33 (EC 2.7.1.-).
GN MKK2 OR SSP33 OR VPL140C OR LPI6C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93233668; PubMed=8386320;
RA Irie K., Takase M., Lee K.S., Levin D.E., Araki H., Matsumoto K.,
RA Oshima Y.;

```

```

"MKK1 and MKK2, which encode Saccharomyces cerevisiae mitogen-
activated protein kinase-kinase homologs, function in the pathway
mediated by protein kinase C".;1993).
RL Mol. Cell. Biol. 13:3076-3083(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313271; PubMed=9168975;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Heblung U., Heumann K., Hilbert H., Hillier L.,
RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
RA Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI. ";
RN Nature 387:103-105(1997).
RN [3]
RP SEQUENCE OF 1-88 FROM N.A.
RC STRAIN=A364A;
RA Mallory M.J., Strich R.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: SERINE/THREONINE PROTEIN KINASE INVOLVED IN A SIGNAL
CC TRANSDUCTION PATHWAY THAT PLAY A ROLE IN YEAST CELL MORPHOGENESIS
CC AND CELL GROWTH. THIS PATHWAY SEEMS TO STARTS BY SMP3; THEN
CC INVOLVES THE KINASE PKC1 THAT MAY ACT ON THE BCK1 KINASE THAT THEN
CC PHOSPHORYLATES MKK1 AND MKK2 WHICH THEMSELVES PHOSPHORYLATE THE
CC MPK1 KINASE.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MAP KINASE KINASE SUBFAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D13785; BAA02933.1; -
CC EMBL; U43703; AAB68220.1; -
CC EMBL; U10280; AAB40938.1; -
CC PIR; S30786; S30786.
CC PIR; B48069; B48069.
CC SGD; S0006061; MKK2.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002290; Ser_thr_kinase.
CC Pfam; PF00069; pkinase; 1.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00220; S_TKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
KW ATP-binding; Phosphorylation.
FT DOMAIN 214 481 PROTEIN KINASE.
FT NP_BIND 220 228 ATP (BY SIMILARITY).
FT BINDING 243 243 ATP (BY SIMILARITY).
FT ACT_SITE 342 342 BY SIMILARITY.
FT MOD_RES 370 370 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 374 374 PHOSPHORYLATION (BY SIMILARITY).
FT CONFLICT 252 252 P -> S (IN REF. 1).
SQ SEQUENCE 506 AA; 56757 MW; 8EACED6C742E148B CRC64;

```

Query Match 0.7%; Score 8; DB 1; Length 506;  
Best Local Similarity 100.0%; Pred. No. 5.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 923 ISERVIGK 930  
|||||||  
DB 314 ISERVIGK 321

## RESULT 6

DXS\_RHOCA STANDARD; PRT; 641 AA.  
AC P26242;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE 1-deoxy-D-xylulose 5-phosphate synthase (EC 4.1.3.37) (1-  
DE deoxyxylulose-5-phosphate synthase) (DXP synthase) (DXPS).  
GN DXS.  
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
OC Rhodobacter.  
OX NCBI\_TaxID=1061;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84259352; PubMed=6744416;  
RA Youvan D.C., Bylina E.J., Alberti M., Begusch H., Hearst J.E.;  
RT "Nucleotide and deduced polypeptide sequences of the photosynthetic  
RT reaction-center, B870 antenna, and flanking polypeptides from R.  
RT capsulata".  
RL Cell 37:949-957(1984).  
CC -!- FUNCTION: Catalyzes the acyloln condensation reaction between C  
CC atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield  
CC 1-deoxy-D-xylulose-5-phosphate (DXP) (By similarity).  
CC -!- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-  
CC deoxy-D-xylulose 5-phosphate + CO(2).  
CC -!- COFACTOR: Thiamine pyrophosphate (By similarity).  
CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; first step.  
CC -!- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first  
CC step.  
CC -!- SUBUNIT: Homodimer (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE TRANSKETOLASE FAMILY. DXS SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; K01183; -; NOT ANNOTATED\_CDS.  
CC DR EMBL; X11165; CAA77557.1; -;  
CC DR PIR; G28771; G28771.  
CC DR InterPro; IPR000360; Transketolase.  
CC DR Pfam; PF02779; transket\_pyr; 1.  
CC DR Pfam; PF02780; transketolase\_C; 1.  
CC DR TIGRfam; TIGR00204; dxs; 1.  
CC DR PROSITE; PS00801; TRANSKETOLASE\_1; 1.  
CC DR PROSITE; PS00802; TRANSKETOLASE\_2; 1.  
CC DR Lyase; Flavo-protein; Thiamine pyrophosphate;  
CC KW Isoprene biosynthesis; Thiamine biosynthesis.  
CC SQ SEQUENCE 641 AA; 67943 MW; CBCC0BF23C012201 CRC64;

Query Match 0.7%; Score 8; DB 1; Length 641;  
Best Local Similarity 100.0%; Pred. No. 7.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 GLTANDIR 472  
|||||||  
DB 615 GLTANDIR 622

## RESULT 7

GREA\_BORBU STANDARD; PRT; 901 AA.  
AC 051157;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Transcription elongation factor grea (Transcript cleavage factor  
DE grea).  
GN GREA OR BB0132.  
OS Borrelia burgdorferi (Lyme disease spirochete).  
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
OX NCBI\_TaxID=139;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=ATCC 35210 / B31;  
RX MEDLINE=98065943; PubMed=9403685;  
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,  
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,  
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,  
RA Peterson J., Kervlavage A.R., Quackenbush J., Salzberg S., Hanson M.,  
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,  
RA Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,  
RA Gariand S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,  
RA Smith H.O., Venter J.C.;  
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia  
RT burgdorferi".  
RL Nature 390:580-586(1997).  
CC -!- FUNCTION: NECESSARY FOR EFFICIENT RNA POLYMERASE TRANSCRIPTION  
CC ELONGATION PAST TEMPLATE-ENCODED ARRESTING SITES. THE ARRESTING  
CC SITES IN DNA HAVE THE PROPERTY OF TRAPPING A CERTAIN FRACTION  
CC OF ELONGATING RNA POLYMERASES THAT PASS THROUGH, RESULTING IN  
CC LOCKED TERNARY COMPLEXES. CLEAVAGE OF THE NASCENT TRANSCRIPT BY  
CC CLEAVAGE FACTORS SUCH AS GREa OR GREB ALLOWS THE RESUMPTION OF  
CC ELONGATION FROM THE NEW 3'-TERMINUS. GREa RELEASES SEQUENCES OF  
CC 2 TO 3 NUCLEOTIDES (BY SIMILARITY).  
CC -!- SIMILARITY: CONTAINS 1 GREa ASSOCIATED DOMAIN 1 (GRAD1).  
CC -!- SIMILARITY: CONTAINS 1 GREa ASSOCIATED DOMAIN 2 (GRAD2).  
CC -!- SIMILARITY: BELONGS TO THE GREa/GREB FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE001124; AAC66506.1; -;  
CC DR HSSP; P21346; IGRJ.  
CC DR TIGR; BB0132; -;  
CC DR InterPro; IPR001437; GreA\_GreB.  
CC DR Pfam; PF01272; GreA\_GreB; 1.  
CC DR Pfam; PF03449; GreA\_GreB\_N; 1.  
CC DR ProDom; PD004918; GreA\_GreB; 1.  
CC DR PROSITE; PS00829; GREAB\_1; 1.  
CC DR PROSITE; PS00830; GREAB\_2; 1.  
CC KW Transcription regulation; DNA-binding; Coiled coil; Complete proteome.  
CC FT DOMAIN 20 170 GRAD1.  
CC FT DOMAIN 171 681  
CC FT DOMAIN 741 901  
CC FT DOMAIN 784 815 COILED COIL (POTENTIAL).  
CC SQ SEQUENCE 901 AA; 105314 MW; FBB81088D735A5F9 CRC64;

Query Match 0.7%; Score 8; DB 1; Length 901;  
Best Local Similarity 100.0%; Pred. No. 9.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 764 DIRVLKAV 771  
|||||||  
DB 359 DIRVLKAV 366

```
RESULT 8
PEX1_HUMAN
ID PEX1_HUMAN STANDARD; PRT; 1283 AA.
AC O43933; Q99994;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Peroxisome biogenesis factor 1 (peroxin-1) (Peroxisome biogenesis
disorder protein 1).
GN PEX1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A., AND VARIANT NALD ASP-843.
RX MEDLINE=98061095; PubMed=9398848;
RA Portsteffen H., Beyer A., Becker E., Epplen C., Pawlak A.,
RA Kunau W.-H., Dödt G.;
RT "Human PEX1 is mutated in complementation group 1 of the peroxisome
biogenesis disorders";
RL Nat. Genet. 17:449-452(1997).
[2]
SEQUENCE FROM N.A., AND VARIANT NALD ASP-843.
RX MEDLINE=98061094; PubMed=9398847;
RA Reuber B.E., Germain-Lee E., Collins C.S., Morrell J.C.,
RA Ameritunga R., Moser H.W., Valle D., Gould S.J.;
RT "Mutations in PEX1 are the most common cause of peroxisome biogenesis
disorders";
RL Nat. Genet. 17:445-448(1997).
[3]
SEQUENCE FROM N.A., AND VARIANTS NALD PRO-664 AND ASP-843.
RX MEDLINE=98208543; PubMed=9539740;
RA Tamura S., Okumoto K., Toyama R., Shimozawa N., Tsukamoto T.,
RA Suzuki Y., Osumi T., Kondo N., Fujiki Y.;
RT "Human PEX1 cloned by functional complementation on a CHO cell mutant
is responsible for peroxisome-deficient Zellweger syndrome of
complementation group I.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:4350-4355(1998).
[4]
SEQUENCE OF 635-1283 FROM N.A.
RA Pauley A.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR STABILITY OF PEX5 AND PROTEIN IMPORT INTO
THE PEROXISOME MATRIX.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- DISEASE: DEFECTS IN PEX1 ARE THE CAUSE OF PEROXISOME BIOGENESIS
DISORDERS (PBD) WHICH ARE LETHAL RECESSIVE DISEASES CAUSED BY
DEFECTS IN PEROXISOME ASSEMBLY. NEONATAL ADRENOLKODYSTROPHY
(NALD) IS ONE OF THE PBD LINKED TO DEFECTS IN PEX1
CC -1- DISEASE: DEFECTS IN PEX1 ARE THE CAUSE OF INFANTILE REFSUM DISEASE
(LRD). LRD IS A FORM OF PHYTANIC ACID STORAGE DISEASE.
CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF030356; AAB99758.1; -
DR EMBL; AF026086; AAB87880.1; -
DR EMBL; AB008112; AAB85162.1; -
DR EMBL; AC000084; AAB46346.1; ALT_SEQ.
DR Genew; HGNC:8850; PEX1.
DR MIM; 602136; -
DR MIM; 202370; -
DR MIM; 266510; -
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_cent.
DR InterPro; IPR003960; AAA_sub.

Query Match 0.7%; Score 8; DB 1; Length 1385;
Best Local Similarity 100.0%; Pred. No. 13; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

PFam; PF00004; AAA; 2.
SMART; SMO0382; AAA; 1.
PROSITE; PS00674; AAA; 1.
Peroxisome; ATP-binding; Repeat; Disease mutation.
NP_BIND 599 606 ATP (POTENTIAL).
NP_BIND 881 888 ATP (POTENTIAL).
L -> P (IN NALD).
FTID=VAR_008876.
VARIANT 664 664 /FTID=VAR_008876.
VARIANT 843 843 G -> D (IN IRD AND NALD).
/FTID=VAR_008877.
SEQUENCE 1283 AA; 142866 MW; 333CE0B15D2E2017 CRC64;

Query Match 0.7%; Score 8; DB 1; Length 1283;
Best Local Similarity 100.0%; Pred. No. 13; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY 68 RLYFGSSY 75
|||||||
DB 1127 RLYFGSSY 1134

RESULT 9
RRPO_PIAMV STANDARD; PRT; 1385 AA.
ID RRPO_PIAMV
AC Q07518;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE RNA replication protein (156 kDa protein) (ORF 1) [Contains: RNA-
directed RNA polymerase (EC 2.7.7.48); Probable helicase].
OS Plantago asiatica mosaic potexvirus (PIAMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.
OX NCBI_TaxID=45416;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=94157481; PubMed=8113746;
RA Solovyev A.G., Novikov V.K., Merits A., Savenkov E.I.,
RA Zelenina D.A., Tyukina L.G., Morozov S.Y.;
RT "Genome characterization and taxonomy of Plantago asiatica mosaic
potexvirus.";
RL J. Gen. Virol. 75:259-267(1994).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=93237851; PubMed=8477237;
RA Solovyev A.G., Novikov V.K., Morozov S.I., Kagramanov V.N.,
RA Atabekov I.G.;
RT "Primary structure of the triple block RNA genes of the Plantago
asiatica mosaic virus.";
RL Dokl. Akad. Nauk SSSR 328:625-628(1993).
CC -1- FUNCTION: RNA-replication. The central part of this protein
possibly functions as an ATP-binding helicase.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
[RNA](N).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Z21647; CAA79761.1; -
DR InterPro; IPR001788; RNA_dep_RNAPol2.
DR InterPro; IPR000606; Viral_helicase1.
DR Pfam; PF00978; RNA_dep_RNAPol2; 1.
DR Pfam; PF01443; Viral_helicase1; 1.
KW ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;
FT NP_BIND 665 672 ATP (POTENTIAL).
SEQUENCE 1385 AA; 155616 MW; 56C177E44678A94F CRC64;

Query Match 0.7%; Score 8; DB 1; Length 1385;
```

```
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 425 ERILSTLR 432
Db 597 ERILSTLR 604

RESULT 10
REV_HV128
ID REV_HV128 STANDARD; PRT; 90 AA.
AC P03869;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE REV protein (Anti-repression transactivator protein) (ART/TRS)
DE (Fragment).
GN REV.
OS Human immunodeficiency virus type 1 (Z-84 isolate) (HIV-1).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11681;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88281278; PubMed=3395517;
RA Yourho J., Josephs S.F., Reitz M.S. Jr., Zagury D., Wong-Staal F.,
RA Gallo R.C.;
RT "Nucleotide sequence analysis of the env gene of a new Zairian
RT isolate of HIV-1.";
RL AIDS Res. Hum. Retroviruses 4:165-173(1988).
CC -!- FUNCTION: REV APPEARS TO ACT POST-TRANSCRIPTIONALLY TO RELIEVE
CC NEGATIVE REPRESSION OF GAG AND ENV PRODUCTION.
CC -!- SUBCELLULAR LOCATION: Nuclear; accumulates in the nucleoli.
CC -!- PTM: PHOSPHOPROTEIN WHOSE STATE OF PHOSPHORYLATION IS MEDIATED
CC BY A SPECIFIC SERINE KINASE ACTIVITY PRESENT IN THE NUCLEUS.
CC -!- MISCELLANEOUS: THE Z-84 ISOLATE WAS TAKEN FROM A 54 YEAR-OLD
CC ZAIREAN MALE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03653; AAA44686.1; -
DR HIV; J03653; REV$JY1.
DR InterPro; IPR000625; REV_protein.
DR Pfam; PF00424; REV; 1
KW Transcription regulation; AIDS; Phosphorylation; Nuclear protein.
FT NON_TER 1
SQ SEQUENCE 90 AA; 10031 MW; B1940D344B0580B7 CRC64;

Query Match 0.6%; Score 7; Length 90;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 425 ERILSTL 431
Db 31 ERILSTL 37

RESULT 11
RL17_CHLTR
ID RL17_CHLTR STANDARD; PRT; 141 AA.
AC P47760; O84514;
DT 01-FEB-1996 (Rel. 33, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L17.
GN RPIQ OR RL17 OR CT506.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
```

```
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=L2/434/Bu; PubMed=7730299;
RX MEDLINE=95247702;
RA Gu L.J., Wenman W.M., Remacha M., Meuser R.U., Coffin J.M., Kaul R.;
RA "Chlamydia trachomatis RNA polymerase alpha subunit: sequence and
RT structural analysis.";
RL J. Bacteriol. 177:2594-2601(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=D/UW-3/Cx; PubMed=97841136;
RX MEDLINE=99000809; PubMed=97841136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
CC -!- SIMILARITY: BELONGS TO THE L17P FAMILY OF RIBOSOMAL PROTEINS.
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT WAS CORRECTED IN POSITION 14.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L39334; AAA74990.1; ALT_INIT.
DR EMBL; AE01323; AAC68107.1; -
DR InterPro; IPR000456; Ribosomal_L17.
DR Pfam; PF01196; Ribosomal_L17; 1
DR ProDom; PD004277; Ribosomal_L17; 1.
DR TIGRams; TIGR00059; L17; 1.
DR PROSITE; PS01167; RIBOSOMAL_L17; 1.
KW Ribosomal protein; Complete proteome.
FT VARIANT 116 116 G->R (IN SEROVAR L2).
SQ SEQUENCE 141 AA; 16152 MW; DAD293A8FB8ED79 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 141;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 926 RVIGKLF 932
Db 98 RVIGKLF 104

RESULT 12
RL17_CHLMU
ID RL17_CHLMU STANDARD; PRT; 142 AA.
AC Q9PJN5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L17.
GN RPIQ OR TC0793.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MoPn / Nigg; PubMed=10684935;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
```



RT pneumoniac AR39"; 28:1397-1406(2000).  
CC -1- SIMILARITY: BELONGS TO THE LI7P FAMILY OF RIBOSOMAL PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE002347; AAF39596.1; -  
DR TIGR; TC0793;  
DR InterPro; IPR000456; Ribosomal\_L17.  
DR Pfam; PF01196; Ribosomal\_L17; 1.  
DR ProDom; PD004277; Ribosomal\_L17; 1.  
DR TIGRFAMs; TIGR00059; L17; 1.  
DR PROSITE; PS01167; RIBOSOMAL\_L17; 1.  
KW Ribosomal protein; Complete proteome.  
SQ SEQUENCE 142 AA; 16223 MW; B17AD1A1760A0767 CRC64;  
  
Query Match 0.6%; Score 7; DB 1; Length 142;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 926 RVIGKLF 932  
Db 98 RVIGKLF 104  
|||||||  
- - - - -  
RESULT 13  
CYNS\_SULTO STANDARD; PRT; 148 AA.  
ID CYNS\_SULTO  
AC Q972WS;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cyanate hydratase (EC 4.2.1.104) (Cyanase) (Cyanate lyase) (Cyanate  
DE hydrolase).  
GN CYNS OR STI023.  
OS Sulfolobus tokodaii.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.  
OX NCBI\_TaxID=111955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JCM 10545 / 7;  
RX MEDLINE=21456156; PubMed=11572479;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S., Y.,  
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,  
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
RA Oshima T., Kikuchi H.;  
RT "Complete genome sequence of an aerobic thermoacidophilic  
RT Crenarchaeon, Sulfolobus tokodaii strain7.";  
RL DNA Res. 8:123-140(2001).  
CC -1- FUNCTION: Catalyzes the reaction of cyanate with bicarbonate to  
CC produce ammonia and carbon dioxide (By similarity).  
CC -1- CATALYTIC ACTIVITY: Cyanate (NCO(-)) + H(2)O = carbamate (H(2)N-  
CC CO-O(-)).  
CC  
CC -1- SIMILARITY: BELONGS TO THE CYANASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AP000984; BAB66048.1; -

DR InterPro; IPR003712; Cyanate\_lyase.  
DR Pfam; PF02560; Cyanate\_lyase; 1.  
DR TIGRFAMs; TIGR00673; cynS; 1.  
KW Lyase; Complete proteome.  
FT ACT\_SITE 89 89 BY SIMILARITY.  
FT ACT\_SITE 92 92 BY SIMILARITY.  
FT ACT\_SITE 115 115 BY SIMILARITY.  
SQ SEQUENCE 148 AA; 17410 MW; 33CB637A45B3BB85 CRC64;  
  
Query Match 0.6%; Score 7; DB 1; Length 148;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 58 ADKIITL 64  
Db 51 ADKIITL 57  
|||||||  
- - - - -  
RESULT 14  
GSHZ\_HELAN STANDARD; PRT; 167 AA.  
ID GSHZ\_HELAN  
AC Q23970;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Glutathione peroxidase 1 (EC 1.11.1.9).  
GN GPXHA-1.  
OS Helianthus annuus (Common sunflower).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;  
OC Heliantheae; Helianthus.  
OX NCBI\_TaxID=4232;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Drevet J.R., Gagne G., Tourvieille de Labrouhe D., Nicolas P.,  
RA Dufau J.P., Ledoigt G., Roedel-Drevet P.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: MAY CONSTITUTE A GLUTATHIONE PEROXIDASE-LIKE  
CC PROTECTIVE SYSTEM AGAINST OXIDATIVE STRESSES.  
CC -1- CATALYTIC ACTIVITY: 2 glutathione + H(2)O(2) = oxidized  
CC glutathione + 2 H(2)O.  
CC -1- SIMILARITY: BELONGS TO THE GLUTATHIONE PEROXIDASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Y14429; CAA74775.1; -  
DR HSSP; P00435; LGPI.  
DR InterPro; IPR000889; Glut\_peroxidase.  
DR Pfam; PF00255; GSHPx; 1.  
DR PRINTS; PR01011; GLUTPROXDASE.  
DR PROSITE; PS00460; GLUTATHIONE\_PEROXID\_1; 1.  
DR PROSITE; PS00763; GLUTATHIONE\_PEROXID\_2; 1.  
KW Oxidoreductase; Peroxidase.  
FT ACT\_SITE 41 41  
SQ SEQUENCE 167 AA; 18842 MW; CFSAC7ACC76558C9 CRC64;  
  
Query Match 0.6%; Score 7; DB 1; Length 167;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 539 VASKCGL 545  
Db 37 VASKCGL 43  
|||||||  
- - - - -  
RESULT 15

GSXK ARATH ID GSXK ARATH STANDARD: PRT: 169 AA.

AC 004922; DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Probable glutathione peroxidase At2g31570 (EC 1.11.1.9).

GS AT2g31570 OR T9H9.9.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;

OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RESULT 16

BTUE\_ECOLI ID BTUE\_ECOLI STANDARD: PRT: 183 AA.

AC P06610;

DT 01-JAN-1988 (Rel. 06, Created)

DT 01-JAN-1988 (Rel. 06, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Vitamin B12 transport periplasmic protein btue.

GN BTUE OR B1710.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI\_TaxID=562;

RP [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=86304184; PubMed=3528129;

RA Friedrich M.J., Deveau L.C., Kadner R.J.;

RA "Nucleotide sequence of the btueD genes involved in vitamin B12

RT transport in Escherichia coli and homology with components of

RT periplasmic-binding-protein-dependent transport systems.;"

RL J. Bacteriol. 167:928-934(1986).

RP [2]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1635;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.;"

RL Science 277:1453-1474 (1997).

RP [3]

RP SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=97251357; PubMed=9097039;

RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,

RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M., Mori T.,

RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Nishio Y.,

RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,

RA Oshima H., Saito N., Sampei G., Seki Y., Sivasubraman S.,

RA Tagami H., Takada J., Takemoto K., Takeuchi Y., Wada C.,

RA Yamamoto Y., Horiuchi T.;

RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome

RT corresponding to the 28.0-40.1 min region on the linkage map.;"

RL DNA Res. 3:363-377(1996).

CC -!- FUNCTION: NOT ESSENTIAL FOR B12 TRANSPORT; HOWEVER, IT IS AN

CC AUXILIARY COMPONENT OF THE TRANSPORT SYSTEM.

CC -!- SUBCELLULAR LOCATION: Periplasmic (Probable).

CC -!- SIMILARITY: BELONGS TO THE GLUTATHIONE PEROXIDASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC -----

CC EMBL; M14031; AAA33527.1; -

CC EMBL; AE000266; AAC74780.1; -

CC EMBL; D90813; BAA15478.1; -

CC EMBL; D90814; BAA15490.1; -

DR EMBL; M14031; AAA33527.1; -

DR EMBL; AE000266; AAC74780.1; -

DR EMBL; D90813; BAA15478.1; -

DR EMBL; D90814; BAA15490.1; -

ACT\_SITE 41 BY SIMILARITY.

GSXK ARATH ID GSXK ARATH STANDARD: PRT: 169 AA.

AC 004922; DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Probable glutathione peroxidase At2g31570 (EC 1.11.1.9).

GS AT2g31570 OR T9H9.9.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;

OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RESULT 16

BTUE\_ECOLI ID BTUE\_ECOLI STANDARD: PRT: 183 AA.

AC P06610;

DT 01-JAN-1988 (Rel. 06, Created)

DT 01-JAN-1988 (Rel. 06, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Vitamin B12 transport periplasmic protein btue.

GN BTUE OR B1710.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI\_TaxID=562;

RP [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=86304184; PubMed=3528129;

RA Friedrich M.J., Deveau L.C., Kadner R.J.;

RA "Nucleotide sequence of the btueD genes involved in vitamin B12

RT transport in Escherichia coli and homology with components of

RT periplasmic-binding-protein-dependent transport systems.;"

RL J. Bacteriol. 167:928-934(1986).

RP [2]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1635;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.;"

RL Science 277:1453-1474 (1997).

RP [3]

RP SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=97251357; PubMed=9097039;

RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,

RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M., Mori T.,

RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Nishio Y.,

RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,

RA Oshima H., Saito N., Sampei G., Seki Y., Sivasubraman S.,

RA Tagami H., Takada J., Takemoto K., Takeuchi Y., Wada C.,

RA Yamamoto Y., Horiuchi T.;

RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome

RT corresponding to the 28.0-40.1 min region on the linkage map.;"

RL DNA Res. 3:363-377(1996).

CC -!- FUNCTION: NOT ESSENTIAL FOR B12 TRANSPORT; HOWEVER, IT IS AN

CC AUXILIARY COMPONENT OF THE TRANSPORT SYSTEM.

CC -!- SUBCELLULAR LOCATION: Periplasmic (Probable).

CC -!- SIMILARITY: BELONGS TO THE GLUTATHIONE PEROXIDASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC -----

CC EMBL; U94495; AAB52725.1; -

CC EMBL; AC007071; AAD24836.1; -

CC EMBL; AY058187; AAL25600.1; -

CC EMBL; AY044330; AAK73271.1; -

CC HSP: P00435; 1GPI.

CC InterPro: IPR000889; Glut\_peroxidase.

CC Pfam: PF00255; GSHPX; 1.

CC PRINTS: PR01011; GLUTPROXDAE.

CC PROSITE: PS00460; GLUTATHIONE\_PEROXID\_1; 1.

CC PROSITE: PS00763; GLUTATHIONE\_PEROXID\_2; 1.

CC Oxidoreductase; Peroxidase.

ACT\_SITE 41 BY SIMILARITY.

DR PIR: E24498; QRECEB.  
 DR HSSP; P00435; 1GP1.  
 DR EcoGene; EG10129; btue.  
 DR InterPro; IPR000889; Glut\_peroxidase.  
 DR Pfam; PF00255; GSHPx; 1.  
 DR PRINTS; PRO1011; GLUTPOXDASE.  
 DR PROSITE; PS00460; GLUTATHIONE\_PEROXID\_1; 1.  
 DR PROSITE; PS00763; GLUTATHIONE\_PEROXID\_2; 1.  
 DR Periplasmic; Cobalt transport; Complete proteome.  
 KW PERIPLASMIC; 183 AA; 20469 MW; C8DB671963A7F235 CRC64;  
 SQ SEQUENCE 183 AA; 20469 MW; C8DB671963A7F235 CRC64;  
 Query Match 0.68; Score 7; DB 1; Length 183;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 539 VASKCGL 545  
 Db 33 VASKCGL 39  
 RESULT 17  
 YQOC\_CAEEL STANDARD; PRT; 137 AA.  
 AC Q09301;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical RNA-binding protein EED8.12 in chromosome II.  
 GN EED8.12.  
 OS Caenorhabditis elegans.  
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
 CC Rhabditidae; Peloderinae; Caenorhabditis.  
 CC NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Chissole S.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
 CC -!- SIMILARITY: STRONG. TO C.ELEGANS EED8.4.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U23484; AAC46772.1; -.  
 DR WormPep; EED8.12; CE01885.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 1.  
 DR SMART; SM00360; RRM; 1.  
 DR PROSITE; PS0102; RRM; 1.  
 DR PROSITE; PS00030; RRM\_RNP\_1; FALSE\_NEG.  
 KW Hypothetical protein; RNA-binding.  
 FT DOMAIN 51 138 RNA-BINDING (RRM).  
 SQ SEQUENCE 137 AA; 21649 MW; AB5C4E7247DD9729 CRC64;  
 Query Match 0.68; Score 7; DB 1; Length 197;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 839 TIEVEE 845  
 Db 73 TIEVEE 79  
 RESULT 18  
 NADD\_BACHD STANDARD; PRT; 207 AA.  
 ID NADD\_BACHD  
 AC Q9KD91;

DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable nicotinate-nucleotide adenyltransferase (EC 2.7.7.18)  
 DE (deamido-NAD(+)-pyrophosphorylase) (deamido-NAD(+) diphosphorylase)  
 DE (Nicotinate mononucleotide adenyltransferase) (NAMN  
 DE adenyltransferase).  
 DE NADD OR BH1326.  
 GN Bacillus halodurans.  
 OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OC NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 CC -!- FUNCTION: CATALYZES THE REVERSIBLE ADENYLATION OF NICOTINATE  
 CC MONONUCLEOTIDE (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: ATP + nicotinate ribonucleotide = diphosphate  
 CC + deamido-NAD(+).  
 CC -!- PATHWAY: NAD biosynthesis.  
 CC -!- SIMILARITY: BELONGS TO THE NADD FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF001511; BAB05045.1; -.  
 DR InterPro; IPR004821; Cyt\_tran\_rel.  
 DR InterPro; IPR004820; Cytidylyltransf.  
 DR InterPro; IPR005248; NAMN\_adntrnsfrase.  
 DR Pfam; PF01467; Cytidylyltransf; 1.  
 DR TIGRFAMs; TIGR00125; cyt\_tran\_rel; 1.  
 DR TIGRFAMs; TIGR00482; TIGR00482; 1.  
 KW Transferase; Nucleotidytransferase; NAD; Complete proteome.  
 SQ SEQUENCE 207 AA; 23434 MW; C2F56B3F85109632 CRC64;  
 Query Match 0.68; Score 7; DB 1; Length 207;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 371 GPEVNV 377  
 Db 152 GPEVNV 158  
 RESULT 19  
 RL13\_CHICK STANDARD; PRT; 210 AA.  
 ID RL13\_CHICK  
 AC P41125;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE 60S ribosomal protein L13 (Breast basic conserved protein 1).  
 GN RPL13 OR BCL1.  
 OS Gallus gallus (Chicken).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;  
 CC Gallus.  
 OC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lens fibers;  
 RX MEDLINE=96437509; PubMed=8840185;

```
RA Savada K., Agata K., Eguchi G.;
RT "Characterization of terminally differentiated cell state by
RT categorizing cDNA clones derived from chicken lens fibers.";
RL Int. J. Dev. Biol. 40:531-535(1996).
CC -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D26318; BAA05377.1; -
DR InterPro; IPR001380; Ribosomal_L13E.
DR Pfam; PF01294; Ribosomal_L13e; 1.
DR ProDom; PD004443; Ribosomal_L13e; 1.
DR PROSITE; PS01104; RIBOSOMAL_L13E; 1.
KW Ribosomal protein.
FT INIT_MET 0 BY SIMILARITY.
SQ SEQUENCE 210 AA; 24264 MW; D8D005B96A8BA9B4 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 210;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 VDPFRRN 304
DB 97 VDPFRRN 103

RESULT 20
RL13_CRIGR STANDARD; PRT; 210 AA.
AC Q92313;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 60S ribosomal protein L13.
GN RPL13.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki Y., Itoh F., Suzuki H., Hinoda Y., Imai K.;
RT "Identification of genes highly expressed in association with G2
RT arrest induced by DNA damage of Chinese hamster ovary cells by
RT differential display.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB014876; BAA34291.1; -
DR InterPro; IPR001380; Ribosomal_L13E.
DR Pfam; PF01294; Ribosomal_L13e; 1.
DR ProDom; PD004443; Ribosomal_L13e; 1.
DR PROSITE; PS01104; RIBOSOMAL_L13E; 1.
KW Ribosomal protein.
FT INIT_MET 0 BY SIMILARITY.
SQ SEQUENCE 210 AA; 24261 MW; 484429F34C565720 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 210;
```

```
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 VDPFRRN 304
DB 97 VDPFRRN 103

RESULT 21
RL13_HUMAN STANDARD; PRT; 210 AA.
ID RL13_HUMAN
AC P26373; Q9BPX0;
DT 01-AUG-1992 (Rel. 23, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L13 (Breast basic conserved protein 1).
GN RPL13 OR BEC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Adams S.M., Helps N.R., Sharp M.G.F., Brammar W.J., Walker R.A.,
RA Varley J.M.;
RT "Isolation and characterization of a novel gene with differential
RT expression in benign and malignant human breast tumours.";
RL Hum. Mol. Genet. 1:91-96(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Cervix, Lung, Lymph, Pancreas, Placenta, and Uterus;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- TISSUE SPECIFICITY: HIGHER LEVELS OF EXPRESSION IN BENIGN BREAST
CC LESIONS THAN IN CARCINOMAS.
CC -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64707; CAA45963.1; -
DR EMBL; BC004954; AAH04954.1; -
DR EMBL; BC007345; AAH07345.1; -
DR EMBL; BC007563; AAH07563.1; -
DR EMBL; BC007805; AAH07805.1; -
DR EMBL; BC010994; AAH10994.1; -
DR EMBL; BC013078; AAH13078.1; -
DR EMBL; BC014167; AAH14167.1; -
DR EMBL; BC020804; AAH20804.1; -
DR EMBL; BC027463; AAH27463.1; -
DR PIR; S23753; S23753
DR Genew; HGNC:10303; RPL13.
DR MIM; 113703; -
DR InterPro; IPR001380; Ribosomal_L13e.
DR Pfam; PF01294; Ribosomal_L13e; 1.
DR ProDom; PD004443; Ribosomal_L13e; 1.
DR PROSITE; PS01104; RIBOSOMAL_L13E; 1.
KW Ribosomal protein.
FT INIT_MET 0 BY SIMILARITY.
FT CONFLICT 111
SQ SEQUENCE 210 AA; 24130 MW; CBBB26F143FAF058 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 210;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 VDPFRRN 304
```

```

Db 97 VDPRRN 103
|||||
RESULT 22
RL13_ICTPU STANDARD; PRT; 210 AA.
AC Q90YV5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 60S ribosomal protein L13.
GN RL13.
OS Ictalurus punctatus (Channel catfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
OC Ictaluridae; Ictalurus.
OX NCBI_TaxID=7998;
RN [1]
RP SEQUENCE FROM N.A.
RA Patterson A.P., Karsi A., Liu Z.J.;
RT "Translational machinery of channel catfish: II. Complementary DNA and
RT expression of the complete set of 47 60S ribosomal proteins."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF401567; AAK95139.1;
DR InterPro; IPR001380; Ribosomal_L13E.
DR Pfam; PF01294; Ribosomal_L13e; 1.
DR ProDom; PD004443; Ribosomal_L13e; 1.
DR PROSITE; PS01104; RIBOSOMAL_L13e; 1.
KW Ribosomal protein.
FT INIT-MET 0 BY SIMILARITY.
SQ SEQUENCE 210 AA; 24254 MW; 5D97B80DD4E64172 CRC64;
Query Match 0.6%; Score 7; DB 1; Length 210;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 298 VDPRRN 304
Db 97 VDPRRN 103
|||||
RESULT 23
RL13_MOUSE STANDARD; PRT; 210 AA.
AC P47963; Q9DCH1; Q9CRZ9;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L13 (A52).
GN RPL13.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Palacios R., Xie X.;
RT Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryonic liver, and Kidney;

```

---

```

RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King S., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzaelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang X.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U28917; AAA69923.1;
DR EMBL; AK002787; BAB23358.1;
DR EMBL; AK010989; BAB27309.1;
DR MGD; MGI:105922; Rpl13.
DR InterPro; IPR001380; Ribosomal_L13E.
DR Pfam; PF01294; Ribosomal_L13e; 1.
DR ProDom; PD004443; Ribosomal_L13e; 1.
DR PROSITE; PS01104; RIBOSOMAL_L13e; 1.
KW Ribosomal protein.
FT INIT-MET 0 BY SIMILARITY.
FT CONFLICT 39 39 Q -> L (IN REF. 1).
FT CONFLICT 189 195 RLFGR -> PTLWQSEQ (IN REF. 1).
FT CONFLICT 202 210 AEQGVKKK -> SEQRCWKRN (IN REF. 1).
SQ SEQUENCE 210 AA; 24174 MW; 3795373F518655F2 CRC64;
Query Match 0.6%; Score 7; DB 1; Length 210;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 298 VDPRRN 304
Db 97 VDPRRN 103
|||||
RESULT 24
RL13_RAT STANDARD; PRT; 210 AA.
AC P41123;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE 60S ribosomal protein L13.
GN RPL13.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=94256964; PubMed=8198561;
RA Olivera J., Wool I.G.;

```

```

RT  The primary structure of rat ribosomal protein L13."
RL  Biochem. Biophys. Res. Commun. 201:102-107(1994).
CC  -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL: X78327; CAA55130.1; -
DR  InterPro: IPR001380; Ribosomal_L13E.
DR  Pfam: PF01294; Ribosomal_L13e; 1.
DR  ProDom: PD004443; Ribosomal_L13e; 1.
DR  ProSITE: PS01104; RIBOSOMAL_L13E; 1.
KW  Ribosomal protein.
FT  INIT_MET 0
FT  CONFLICT 55 55 R -> S (IN CDNA).
SQ  SEQUENCE 210 AA; 24178 MW; 379A6737F30C6EE9 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 210;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 VDPRRRN 304
DB 97 VDPRRRN 103
|||||

RESULT 25
YDHL_HSV57
ID YDHL_HSV57 STANDARD; PRT; 214 AA.
AC P25049.
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 24.1 kDa protein in DHFR 3'region (ORF2).
OS Herpesvirus saimiri (strain 484-77).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10382;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90279084; PubMed=2161952;
RA Geck P., Whitaker S.A., Medveczky M.M., Medveczky P.G.;
RT "Expression of collagenlike sequences by a tumor virus, herpesvirus
RL J. Virol. 64:3509-3515(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M31964; AAA46153.1; -
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 214 AA; 24142 MW; 9869BDBDA89FA42A CRC64;

Query Match 0.6%; Score 7; DB 1; Length 214;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 KSFLEKY 26
DB 107 KSFLEKY 113
|||||

RESULT 26

```

```

RL13_SPOFR STANDARD; PRT; 219 AA.
ID RL13_SPOFR
AC Q962U1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L13.
GN RPL13.
OS Spodoptera frugiperda (Fall armyworm).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Ditrysia; Noctuoidea; Noctuidae; Amphipyrinae; Spodoptera.
OX NCBI_TaxID=7108;
RN [1]
RP SEQUENCE FROM N.A.
RA Landais I., Ogliastro M., Mita K., Nohata J., Lopez-Ferber M.,
RA Duonor-Cerutti M., Fournier P., Devauchelle G.;
RT "Full-length ribosomal protein sequence from an EST library of
RT Spodoptera frugiperda cells (Sf9).";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF400183; AAK92155.1; -
DR InterPro: IPR001380; Ribosomal_L13E.
DR Pfam: PF01294; Ribosomal_L13e; 1.
DR ProDom: PD004443; Ribosomal_L13e; 1.
DR ProSITE: PS01104; RIBOSOMAL_L13E; 1.
KW Ribosomal protein.
SQ SEQUENCE 219 AA; 24963 MW; 9FF918EB07CA3449 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 219;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 VDPRRRN 304
DB 97 VDPRRRN 103
|||||

RESULT 27
GTCL_RAT
ID GTCL_RAT STANDARD; PRT; 220 AA.
AC P04904;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glutathione S-transferase Yc-1 (EC 2.5.1.18) (Chain 2) (GST Yc1)
DE (GST class-alpha).
GN GSTY1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fischer 344; TISSUE=Liver;
RX MEDLINE=94327654; PubMed=8051171;
RA Hayes J.D., Nguyen T., Judah D.J., Petersson D.G., Neal G.E.;
RT "Cloning of cDNAs from fetal rat liver encoding glutathione S-
RT transferase Yc polypeptides. The Yc2 subunit is expressed in adult
RT rat liver resistant to the hepatocarcinogen aflatoxin B1.";
RL J. Biol. Chem. 269:20707-20717(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85182743; PubMed=2985614;

```

RA Telakowski-Hopkins C.A., Rodkey K.A., Bennett C.D., Lu A.Y.H.,  
 RA Pickett C.B.;  
 RT "Rat liver glutathione S-transferases. Construction of a cDNA clone  
 RT complementary to a Yc mRNA and prediction of the complete amino acid  
 RT sequence of a Yc subunit.";  
 RN J. Biol. Chem. 260:5820-5825(1985).  
 RN [3]  
 RN SEQUENCE OF 74-220 FROM N.A.  
 RP MEDLINE=84264589; PubMed=6204982;  
 RA Tu C.-P.D., Lai H.-C.J., Li N.-O., Weiss M.J., Reddy C.C.;  
 RA "The Yc and Ya subunits of rat liver glutathione S-transferases are  
 RT the products of separate genes.";  
 RL J. Biol. Chem. 259:9434-9439(1984).  
 RN [4]  
 RN PARTIAL SEQUENCE.  
 RP STRAIN=Fischer 344; TISSUE=Liver;  
 RX MEDLINE=92061940; PubMed=1953636;  
 RA Hayes J.D., Judah D.J., McLellan L.I., Kerr L.A., Peacock S.D.,  
 RA Neal G.E.;  
 RT "Ethoxyquin-induced resistance to aflatoxin B1 in the rat is  
 RT associated with the expression of a novel alpha-class glutathione  
 RT S-transferase subunit, Yc2, which possesses high catalytic activity  
 RT for aflatoxin B1-8,9-epoxide.";  
 RL Biochem. J. 279:385-398(1991).  
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER  
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES. HAS  
 CC SUBSTANTIAL ACTIVITY TOWARD AFLATOXIN B1-8,9-EPOXIDE.  
 CC -1- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.  
 CC -1- SUBUNIT: HETERODIMER OF YC1 AND YC2.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- DEVELOPMENTAL STAGE: LIVER FROM ADULT FEMALE RATS CONTAINS ABOUT  
 CC 2-FOLD GREATER LEVELS OF YC1 THAN IS FOUND IN LIVER FROM ADULT  
 CC MALE RATS.  
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X78848; CAA55405.1; -;  
 DR EMBL; K01932; AAA41294.1; -;  
 DR EMBL; S72505; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; A26753; A26753.  
 DR PIR; A26754; A26754.  
 DR HSSP; P08263; IGSD.  
 DR InterPro; IPR004046; GST\_Cterm.  
 DR InterPro; IPR004045; GST\_Nterm.  
 DR Pfam; PF00043; GST\_C; 1.  
 DR Pfam; PF02798; GST\_N; 1.  
 KW Transferase; Multigene family.  
 FT INIT\_MET 0  
 FT CONFLICT 101 101 L -> I (IN REF. 2).  
 FT CONFLICT 183 183 L -> K (IN REF. 2).  
 SQ SEQUENCE 220 AA; 25188 MW; F40F5FC0757EE93F CRC64;  
 Query Match 0.6%; Score 7; DB 1; Length 220;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1089 KDKARNR 1095  
 Db 124 KDKARNR 130  
 RESULT 28  
 GTC2\_RAT  
 ID GTC2\_RAT STANDARD; PRT; 220 AA.  
 AC P46418;  
 DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glutathione S-transferase Yc-2 (EC 2.5.1.18) (Chain 2) (GST Yc2)  
 DE (GST class-alpha).  
 GN GSTYC2 OR GSTA5.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Fischer 344; TISSUE=Liver;  
 RX MEDLINE=94327654; PubMed=8051171;  
 RA Hayes J.D., Nguyen T., Judah D.J., Petersson D.G., Neal G.E.;  
 RT "Cloning of cDNAs from fetal rat liver encoding glutathione S-  
 RT transferase Yc polypeptides. The Yc2 subunit is expressed in adult  
 RT rat liver resistant to the hepatocarcinogen aflatoxin B1.";  
 RL J. Biol. Chem. 269:20707-20717(1994).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=96358489; PubMed=8761455;  
 RA Fulford D.J., Hayes J.D.;  
 RT "Characterization of the rat glutathione S-transferase Yc2 subunit  
 RT gene, GSTA5: identification of a putative antioxidant-responsive  
 RT element in the 5'-flanking region of rat GSTA5 that may mediate  
 RT chemoprotection against aflatoxin B1.";  
 RL Biochem. J. 318:75-84(1996).  
 RN [3]  
 RN PARTIAL SEQUENCE.  
 RP STRAIN=Fischer 344; TISSUE=Liver;  
 RX MEDLINE=92061940; PubMed=1953636;  
 RA Hayes J.D., Judah D.J., McLellan L.I., Kerr L.A., Peacock S.D.,  
 RA Neal G.E.;  
 RT "Ethoxyquin-induced resistance to aflatoxin B1 in the rat is  
 RT associated with the expression of a novel alpha-class glutathione  
 RT S-transferase subunit, Yc2, which possesses high catalytic activity  
 RT for aflatoxin B1-8,9-epoxide.";  
 RL Biochem. J. 279:385-398(1991).  
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER  
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES. HAS  
 CC SUBSTANTIAL ACTIVITY TOWARD AFLATOXIN B1-8,9-EPOXIDE.  
 CC -1- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.  
 CC -1- SUBUNIT: HETERODIMER OF YC1 AND YC2.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: LIVER, NASAL MUCOSA AND EPIDIDYMIS.  
 CC -1- DEVELOPMENTAL STAGE: LIVER FROM ADULT FEMALE RATS CONTAINS ABOUT  
 CC 10-FOLD GREATER LEVELS OF YC2 THAN IS FOUND IN LIVER FROM ADULT  
 CC MALE RATS.  
 CC -1- INDUCTION: BY ETHOXYQUIN, OLTIPRAZ, BUTYLATED HYDROXYANISOLE, AND  
 CC PHENOBARBITOL.  
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X78847; CAA55404.1; -;  
 DR EMBL; S72506; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; S82820; AAB46796.1; -;  
 DR HSSP; P08263; IGSD.  
 DR InterPro; IPR004046; GST\_Cterm.  
 DR InterPro; IPR004045; GST\_Nterm.  
 DR Pfam; PF00043; GST\_C; 1.  
 DR Pfam; PF02798; GST\_N; 1.  
 KW Transferase; Multigene family.  
 FT INIT\_MET 0  
 FT CONFLICT 101 101 L -> I (IN REF. 2).  
 FT CONFLICT 183 183 L -> K (IN REF. 2).  
 SQ SEQUENCE 220 AA; 25216 MW; DE8D9D43365664F5 CRC64;  
 Query Match 0.6%; Score 7; DB 1; Length 220;

```
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
DB 124 KDKARNR 130
|||||

RESULT 29
KPYK_THIELI
ID KPYK_THIELI STANDARD; PRT; 220 AA.
AC Q56301;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyruvate kinase (EC 2.7.1.40) (PK) (fragment).
GN PKI.
OS Thermococcus litoralis.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Thermococcus.
OX NCBI_TaxID=2265;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 5475;
RX MEDLINE=95370164; PubMed=7642512;
RA Kletzin A., Mukund S., Kelley-Crouse T.L., Chan M.K., Rees D.C.,
RA Adams M.W.W.;
RT "Molecular characterization of the genes encoding the tungsten-
RT containing aldehyde ferredoxin oxidoreductase from Pyrococcus
RT furiosus and formaldehyde ferredoxin oxidoreductase from Thermococcus
RT litoralis.";
RL J. Bacteriol. 177:4817-4819(1995).
CC -!- CATALYTIC ACTIVITY: ATP + pyruvate = ADP + phosphoenolpyruvate.
CC -!- PATHWAY: GLYCOLYSIS; final step.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE PYRUVATE KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X83963; CAA58793.1; -
CC HSP; P11974; LPKN.
CC InterPro: IPR001697; Pyruvate_kinase.
CC Pfam: PF00224; PK; 1.
CC ProDom: PD001009; Pyruvate_kinase; 1.
CC PROSITE: PS00110; PYRUVATE_KINASE; PARTIAL.
KW Transferase; Kinase; Glycolysis; Magnesium.
FT NON_TER 220 220
SQ SEQUENCE 220 AA; 24338 MW; A60B906BC742FCRA CRC64;

Query Match 0.6%; Score 7; DB 1; Length 220;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 VKSFLEK 25
DB 209 VKSFLEK 215
|||||

RESULT 30
EUTQ_SALTY
ID EUTQ_SALTY STANDARD; PRT; 229 AA.
AC Q9ZFEV5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ethanolamine utilization protein eutQ.
GN EUTQ OR STM2468.

Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 VKSFLEK 25
DB 209 VKSFLEK 215
|||||

RESULT 31
RADCBACSU
ID RADCBACSU STANDARD; PRT; 231 AA.
AC Q02170;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA repair protein radC homolog.
GN RADC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93015731; PubMed=1400224;
RA Levin P.A., Margolis P.S., Setlow P., Losick R., Sun D.;
RT "Identification of Bacillus subtilis genes for septum placement and
RT shape determination.";
RL J. Bacteriol. 174:6717-6728(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=93259962; PubMed=8387996;
RA Butler Y.X., Abhayawardhane Y., Stewart G.C.;
```



\*Amplification of the Bacillus subtilis maf gene results in arrested septum formation.\*  
 J. Bacteriol. 175:3139-3145(1993).  
 [3]  
 SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Arvedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S., Borris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Connerston I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T., Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G., Giuseppe G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klier-Bianchard M., Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G., Rev M., Reynolds S., Rieger M., Rivoita C., Rocheau E., Roche B., Rose M., Sadaie I., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A., Tsato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis";  
 RL Nature 390:249-256(1997).  
 CC -1- FUNCTION: INVOLVED IN DNA REPAIR (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE RADC FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M96343; AAA22396.1; -  
 DR EMBL; L08793; AAA22583.1; -  
 DR EMBL; Z99118; CAB14764.1; -  
 DR PIR; B45239; B45239.  
 DR Subtilist; B010325; radc.  
 DR InterPro; IPR003583; HHH\_1.  
 DR InterPro; IPR000445; HHH.  
 DR InterPro; IPR001405; RadC.  
 DR Pfam; PF00633; HHH; 1.  
 DR ProDom; PD007415; RadC; 1.  
 DR SMART; SMO0278; HHH; 1.  
 DR TIGRFAMs; TIGR00608; radc; 1.  
 DR PROSITE; PS01302; RADC; 1.  
 KW DNA repair; Complete proteome.  
 SQ SEQUENCE 231 AA; 26146 MW; B93FC0528C5B49F4 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 231;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 VIGDKKF 442  
 |||||  
 Db 216 VIGDKKF 222

## RESULT 32

1431\_ENTHI  
 ID 1431\_ENTHI STANDARD; PRT; 239 AA.  
 AC P42848;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE 14-3-3 protein 1 (14-3-3-1).  
 OS Entamoeba histolytica.  
 OC Eukaryota; Entamoebidae; Entamoeba.  
 OX NCBI\_TaxID=5759;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HM-1;IMSS;  
 RA Samuelson J., Shen P., Meckler G., Descoteaux S., Fu H., Lohia A.;  
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE 14-3-3 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U13418; AAA80185.1; -  
 DR EMBL; P29312; 1A38.  
 DR InterPro; IPR000308; 14-3-3.  
 DR Pfam; PF00244; 14-3-3; 1.  
 DR PRINTS; PR00305; 1433ZETA.  
 DR ProDom; PD000600; 14-3-3; 1.  
 DR SMART; SMO0103; 14\_3\_3; 1.  
 DR PROSITE; PS00796; 1433\_1; 1.  
 DR PROSITE; PS00797; 1433\_2; 1.  
 KW Multigene family.  
 SQ SEQUENCE 239 AA; 27338 MW; 6BDE1496C8428FFC CRC64;

Query Match 0.6%; Score 7; DB 1; Length 239;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 409 LFPKAST 415  
 |||||  
 Db 108 LLPKAST 114

## RESULT 33

CCS\_MOUSE  
 ID CCS\_MOUSE STANDARD; PRT; 274 AA.  
 AC Q9WU84; Q9CRJ9;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Copper chaperone for superoxide dismutase (Superoxide dismutase copper chaperone).  
 DE CCS OR CCSO.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Bartnikas T.B., Waggoner D.J., Gitlin J.D.;  
 RT "The mouse copper chaperone for superoxide dismutase (CCS).";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NOD; TISSUE=Brain;  
 RX MEDLINE=20237523; PubMed=10773661;  
 RA Moore S.D., Chen M.M., Cox D.W.;

RT "Cloning and mapping of murine superoxide dismutase copper chaperone  
 RL (ccsd) and mapping of the human ortholog.";  
 RN Cytogenet. Cell Genet. 88:35-37(2000).  
 [3]

RP SEQUENCE OF 103-274 FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryonic stem cells;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).

CC -1- FUNCTION: Delivers copper to copper zinc superoxide dismutase

CC (SOD1).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -1- TISSUE SPECIFICITY: Ubiquitous.

CC -1- SIMILARITY: CONTAINS 1 HMA DOMAIN.

CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE CU-ZN

CC SUPEROXIDE DISMUTASE FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
 CC -----

DR EMBL; AF121906; AAD23632.1; -

DR EMBL; AF173379; AAF70242.1; -

DR EMBL; AK010264; BAB26806.1; -

DR HSSP; P15107; IXSO.

DR MGD; MGI:1333783; Ccsd.

DR InterPro; IPR001934; HeavyMe.transpt.

DR InterPro; IPR001424; SOD\_CU\_ZN.

DR Pfam; PF00080; sodecu; 1.

DR Pfam; PF00403; HMA; 1.

DR PRINTS; PR00068; CUZNDISMUTASE.

DR PRODOM; PD000469; SOD\_CU\_ZN; 1.

DR PROSITE; PS01047; HMA\_1; FALSE\_NEG.

DR PROSITE; PS00846; HMA\_2; 1.

DR PROSITE; PS00087; SOD\_CU\_ZN\_1; FALSE\_NEG.

DR PROSITE; PS00332; SOD\_CU\_ZN\_2; 1.

KW Chaperone; Copper; zinc; Metal-binding.

FT DOMAIN 12 75

FT HMA.

FT DOMAIN 88 234 SUPEROXIDE DISMUTASE-LIKE.

FT METAL 22 22 COPPER (POTENTIAL).

FT METAL 25 25 COPPER (POTENTIAL).

FT METAL 130 130 COPPER (BY SIMILARITY).

FT METAL 132 132 COPPER (BY SIMILARITY).

FT METAL 147 147 COPPER AND ZINC (BY SIMILARITY).

FT METAL 155 155 ZINC (BY SIMILARITY).

FT METAL 164 164 ZINC (BY SIMILARITY).

FT METAL 167 167 ZINC (BY SIMILARITY).

FT METAL 201 201 COPPER (BY SIMILARITY).

SQ SEQUENCE 274 AA; 28911 MW; 19DCE48376C9D5A2 CRC64;

Query Match

0.6%; Score 7; DB 1; Length 274;

Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 GTVCALE 36

Db 9 GTVCALE 15

RESULT 34

UPK\_BACSU

ID UPK\_BACSU STANDARD; PRT; 276 AA.

AC P94507;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Putative undecaprenol kinase (EC 2.7.1.66) (Bacitracin resistance

DE protein).

GN UPK OR BACA.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=168;

RX MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,

RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,

RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,

RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,

RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,

RA Denton K.D., Errington J., Fabret C., Ferrari E., Foulger D.,

RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,

RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,

RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,

RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,

RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,

RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,

RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,

RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,

RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,

RA Parto V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,

RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,

RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,

RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,

RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,

RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,

RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,

RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,

RA Winters P., Wiputt A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,

RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,

RT "The complete genome sequence of the Gram-positive bacterium Bacillus

RT subtilis.";

RL Nature 390:249-256(1997).

RN [2]

RP SEQUENCE OF 31-276 FROM N.A.

RC STRAIN=PB131;

RA De Rossi E.;

RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: Probably phosphorylates undecaprenol to undecaprenyl

CC phosphate. Confers resistance to bacitracin (By similarity).

CC -1- CATALYTIC ACTIVITY: ATP + undecaprenol -> ADP + undecaprenyl

CC phosphate.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).

CC -1- MISCELLANEOUS: Bacitracin is thought to be involved in inhibition

CC of peptidoglycan synthesis by sequestering undecaprenyl

CC diphosphate reducing the pool of lipid carrier available.

CC -1- SIMILARITY: BELONGS TO THE UPK FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; Z99119; CAB15093.1; -  
DR EMBL; U87792; AAB47703.1; ALT\_INIT.  
DR Subtilist; BG13951; upk.  
DR InterPro; IPR003824; BACA.  
DR Pfam; PF02673; BACA; 1.  
DR TIGRFAMs; TIGR00753; undec\_kin\_baca; 1.  
KW Transferase; Kinase; Antibiotic resistance; Transmembrane;  
FT TRANSMEM 49 58 POTENTIAL.  
FT TRANSMEM 89 111 POTENTIAL.  
FT TRANSMEM 117 139 POTENTIAL.  
FT TRANSMEM 194 216 POTENTIAL.  
FT TRANSMEM 226 248 POTENTIAL.  
FT TRANSMEM 255 274 POTENTIAL.  
SQ SEQUENCE 276 AA; 30345 MW; 2ED137A85F5833BD CRC64;  
  
Query Match 0.6%; Score 7; DB 1; Length 276;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 966 FEDYIDE 972  
|||||||  
DB 111 FEDYIDE 117  
  
RESULT 35  
Y620\_METJA STANDARD; PRT; 291 AA.  
ID Y620\_METJA  
AC Q58037;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein MJ0620.  
GN MJ0620.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2651 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kervlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
jannaschii.";  
RL Science 273:1058-1073(1996).  
CC -1- SIMILARITY: TO M.JANNASCHII MJ1001 AND TO E.COLI RIBOSOMAL PROTEIN  
CC S6 MODIFICATION PROTEIN (RIMK).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; U67510; AAB98615.1; -  
DR TIGR; MJ0620; -  
DR InterPro; IPR004666; RimK\_fam.

DR TIGRFAMs; TIGR00768; rimk\_fam; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 291 AA; 33277 MW; 2B6A3438290A3B0E CRC64;  
  
Query Match 0.6%; Score 7; DB 1; Length 291;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 204 PSWIGLS 210  
|||||||  
DB 264 PSWIGLS 270  
  
RESULT 36  
UL07\_HSVEB STANDARD; PRT; 303 AA.  
ID UL07\_HSVEB  
AC P28945;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE Gene 55 protein.  
GN 55.  
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Varicellovirus.  
OX NCBI\_TaxID=31520;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92295566; PubMed=1318606;  
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;  
RT "The DNA sequence of equine herpesvirus-1.";  
RL Virology 189:304-316(1992).  
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL7,  
CC EHV-1 55, VZV 53, EBV BBRF2, HCMV UL103 AND HVS-1 42.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M86664; AAB02490.1; -  
DR PIR; A36801; WZBEE7.  
DR InterPro; IPR002600; Herpes\_UL7.  
DR Pfam; PF01677; Herpes\_UL7; 1.  
SQ SEQUENCE 303 AA; 33854 MW; 4449A86C70ACA0ED CRC64;  
  
Query Match 0.6%; Score 7; DB 1; Length 303;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1107 LSHRLVL 1113  
|||||||  
DB 130 LSHRLVL 136  
  
RESULT 37  
GLSA\_STRCO STANDARD; PRT; 307 AA.  
ID GLSA\_STRCO  
AC P57755;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Probable glutaminase SCO7049 (EC 3.5.1.2).  
GN SCO7049 OR S4G1.15.  
OS Streptomyces coelicolor.  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=A3(2) / M145;  
RX MEDLINE=21996410; PubMed=12000953;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorrek A., Woodward J., Barrrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
RT coelicolor A3(2).";  
RL Nature 417:141-147(2002).  
CC -!- CATALYTIC ACTIVITY: L-glutamine + H(2)O -> L-glutamate + NH(3).  
CC -!- SIMILARITY: BELONGS TO THE GLUTAMINASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AL391039; CAC01547.1.;  
KW Hydrolase; Complete proteome.  
SQ SEQUENCE 307 AA; 32555 MW; CCF51DFAE88D2079 CRC64;  
  
Query Match 0.6%; Score 7; DB 1; Length 307;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 533 GDFAHV 539  
DB 247 GDFAHV 253  
|||||||  
  
RESULT 38  
LIP\_PSESP STANDARD; PRT; 311 AA.  
AC P26877;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Lactonizing lipase precursor (EC 3.1.1.3) (Triacylglycerol lipase).  
GN Pseudomonas sp. (strain 109).  
OS Bacteria; Proteobacteria.  
OX NCBI\_TaxID=306;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 27-45.  
RX MEDLINE=92011544; PubMed=1917947;  
RA Ihara F., Kageyama Y., Hirata M., Nihira T., Yamada Y.;  
RT "Purification, characterization, and molecular cloning of lactonizing  
RT lipase from Pseudomonas species.";  
RL J. Biol. Chem. 266:18135-18140(1991).  
CC -!- FUNCTION: CATALYZES THE SYNTHESIS OF MACROCYCLIC LACTONES IN  
CC ANHYDROUS ORGANIC SOLVENTS.  
CC -!- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O -> diacylglycerol + a  
CC fatty acid anion.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: BELONGS TO THE AB HYDROLASE SUPERFAMILY. PSEUDOMONAS  
CC LIPASE FAMILY.  
CC -!- SIMILARITY: BELONGS TO THE ABHYDROLASE SUPERFAMILY. LIPASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

CC EMBL; D10166; BAA01035.1; -.  
DR PIR; A40943; A40943.  
DR HSP; P22088; 3LIP.  
DR InterPro: IPR000073; Abhydrolase.  
DR InterPro: IPR000734; Lipase.  
DR InterPro: IPR000379; Ser\_estrs\_site.  
DR Pfam: PF00381; abhydrolase; 1.  
DR PROSITE: PS00120; LIPASE\_SER; 1.  
KW Hydrolase; Lipid degradation; Signal.  
FT SIGNAL 1 26  
FT CHAIN 27 311 LACTONIZING LIPASE.  
FT ACT\_SITE 108 108 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 235 235 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 277 277 CHARGE RELAY SYSTEM (BY SIMILARITY).  
SQ SEQUENCE 311 AA; 32737 MW; 27AC2F3DD3B334D1 CRC64;  
  
Query Match 0.6%; Score 7; DB 1; Length 311;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 133 FLSSGSA 139  
DB 170 FLSSGSA 176  
|||||||  
  
RESULT 39  
ISPH\_BACHD STANDARD; PRT; 314 AA.  
ID ISPH\_BACHD  
AC Q9KDS7;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE ISPH protein.  
GN ISPH OR LYTB OR BH1382.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
CC -!- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
CC into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate  
CC (DMAPP) (By similarity).  
CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh  
CC (last) step.  
CC -!- SIMILARITY: BELONGS TO THE ISPH FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AP001511; BAB05101.1.;  
DR InterPro: IPR003451; LytB.  
DR Pfam: PF02401; LytB; 1.  
DR TIGRFAMs: TIGR00216; lytB; 1.  
KW Isoprene biosynthesis; Complete proteome.  
SQ SEQUENCE 314 AA; 34841 MW; BC0C12493B22698F CRC64;  
  
Query Match 0.6%; Score 7; DB 1; Length 314;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 56 ISADKII 62
Db 159 ISADKII 165

RESULT 40
VG2_SPV4
ID VG2_SPV4 STANDARD; PRT; 320 AA.
AC P11334;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-JUL-1989 (Rel. 11, Last annotation update)
DE Gene 2 protein.
GN 2.
OS Spiroplasma virus 4 (SPV4).
OC Viruses; ssDNA viruses; Microviridae; Spiromicrovirus.
OX NCBI_TaxID=10855;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80032809; PubMed=2822658;
RA Renaudin J., Pascarel M.-C., Bove J.-M.;
RT "Spiroplasma virus 4: nucleotide sequence of the viral DNA,
RT regulatory signals, and proposed genome organization.";
RL J. Bacteriol. 169:4950-4961(1987).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17988; -; NOT_ANNOTATED_CDS.
DR PIR; H29825; G2BPSV.
SQ SEQUENCE 320 AA; 38090 MW; B4C1B2BCACIA2058 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 320;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 849 NYIVNDS 855
Db 36 NYIVNDS 42

RESULT 41
Y526_CHLPN
ID Y526_CHLPN STANDARD; PRT; 329 AA.
AC Q92826; Q9JQ87;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein CP0526/CP0226/CPJ0526.
GN CPN0526 OR CP0226 OR CPJ0526
OS Chlamydia pneumoniae (Chlamydia phila pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206506; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,

```

---

```

RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- SIMILARITY: BELONGS TO THE SIS FAMILY. GUTQ/KPSF SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001637; AAD18666.1; -;
DR EMBL; AE002183; AAF38092.1; -;
DR EMBL; AF002547; BAA98732.1; -;
DR TIGR; CP0226; -;
DR InterPro; IPR000644; CBS_domain.
DR InterPro; IPR004800; KpsF.
DR InterPro; IPR001347; SIS.
DR Pfam; PF00571; CBS; 2.
DR Pfam; PF01380; SIS; 1.
DR SMART; SM00116; CBS; 1.
DR TIGRFAMs; TIGR00393; kpsF; 1.
KW Hypothetical protein; ATP-binding; Repeat; CBS domain;
KW Complete proteome.
FT NP_BIND 52 57 ATP (POTENTIAL).
FT DOMAIN 204 258 CBS 1.
FT DOMAIN 272 327 CBS 2.
FT SEQUENCE 329 AA; 35650 MW; 27A3023B6ABE6D2F CRC64;

Query Match 0.6%; Score 7; DB 1; Length 329;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 623 KQKEAYD 629
Db 17 KQKEAYD 23

RESULT 42
PDXA_FUSNN
ID PDXA_FUSNN STANDARD; PRT; 332 AA.
AC Q8RGU0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 4-hydroxythreonine-4-phosphate dehydrogenase (EC 1.1.1.262) (4-
DE (phosphohydroxy)-L-threonine dehydrogenase).
GN PDXA OR FN0226.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,

```

```

RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyprides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
CC
CC -!- FUNCTION: Catalyzes the NAD-dependent oxidation of 4-
CC (phosphohydroxy)-L-threonine (HTP) into 2-amino-3-oxo-4-
CC form 1-amino-3-(phosphohydroxy)propan-2-one (3-amino-2-oxopropyl
CC phosphate) (by similarity).
CC
CC -!- CATALYTIC ACTIVITY: 4-(phosphonoxy)-threonine + NAD(+) = 2-
CC amino-3-oxo-4-phosphonoxybutyrate + NADH
CC
CC -!- PATHWAY: De novo synthesis of pyridoxine (vitamin B6) and
CC pyridoxal phosphate.
CC
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC
CC -!- SIMILARITY: BELONGS TO THE PDXA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AE010535; AAL94432.1; -
CC KW Pyridoxine biosynthesis; Oxidoreductase; NAD; Complete proteome.
CC SQ SEQUENCE 332 AA; 36547 MW; A9032E3F6E7081FC CRC64;
CC
CC Query Match 0.6%; Score 7; DB 1; Length 332;
CC Best Local Similarity 100.0%; Pred. No. 49;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 711 TLEYGVV 717
CC |
CC Db 85 TLEYGVV 91
CC
CC RESULT 43
CC RPOA_PSEAE
CC ID RPOA_PSEAE STANDARD; PRT; 333 AA.
CC AC 052760;
CC DT 30-MAY-2000 (Rel. 39, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (Transcriptase
CC alpha chain) (RNA polymerase alpha subunit).
CC GN RPOA OR PA4238.
CC OS Pseudomonas aeruginosa.
CC OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
CC OC Pseudomonas.
CC OX NCBI_TaxID=287;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=FR01;
CC RX MEDLINE=99296583; PubMed=10368149;
CC RA Ma J.-F., Ochsner U.A., Klotz M.G., Nanayakkara V.K., Howell M.L.,
CC RA Johnson Z., Posey J.E., Vasili M.L., Moraco J.J., Hassett D.J.;
CC RT "Bacterioferritin A modulates catalase A (KatA) activity and
CC resistance to hydrogen peroxide in Pseudomonas aeruginosa.";
CC RL J. Bacteriol. 181:3730-3742(1999).
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=ATCC 15692 / PA01;
CC RX MEDLINE=20437337; PubMed=10984043;
CC RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
CC RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
CC RA Garber R.L., Goitry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
CC RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
CC RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
CC RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
CC RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
CC opportunistic pathogen.";

```

```

RL Nature 406:959-964(2000).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC (RNA)(N).
CC
CC -!- SUBUNIT: CONSISTS OF A SIGMA FACTOR AND THE RNAP CORE ENZYME WHICH
CC IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, 1 BETA' CHAIN AND 1
CC OMEGA CHAIN (BY SIMILARITY).
CC
CC -!- DOMAIN: THE AMINO-TERMINAL PORTION IS INVOLVED IN THE ASSEMBLY OF
CC CORE RNAP, WHEREAS THE C-TERMINAL IS INVOLVED IN INTERACTION WITH
CC TRANSCRIPTIONAL REGULATORS (BY SIMILARITY).
CC
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE ALPHA CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF047025; AAC03116.1; -
CC DR EMBL; AE004841; AAG07626.1; -
CC DR HSSP; P00574; 1COO.
CC DR InterPro; IPR001700; RNA_pol_A_bac.
CC DR Pfam; PF01000; RNA_pol_A_bac; 1.
CC DR Pfam; PF03118; RNA_pol_A_CTD; 1.
CC DR ProDom; PD001179; RNA_pol_A_bac; 1.
CC KW Transference; Transcription; DNA-directed RNA polymerase;
CC KW Complete proteome.
CC FT CONFLICT 326 333 KKDKATA -> TERROGHCLIVVITERKVKWKGIEPCAIVK
CC VVVT (IN REF. 1).
CC SQ SEQUENCE 333 AA; 36649 MW; 50706D2926207CA9 CRC64;
CC
CC Query Match 0.6%; Score 7; DB 1; Length 333;
CC Best Local Similarity 100.0%; Pred. No. 49;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 832 IQLDHDV 838
CC |
CC Db 115 IQLDHDV 121
CC
CC RESULT 44
CC Y094_CAEEL
CC ID Y094_CAEEL STANDARD; PRT; 335 AA.
CC AC P41844;
CC DT 01-NOV-1995 (Rel. 32, Created)
CC DT 01-NOV-1995 (Rel. 32, Last sequence update)
CC DT 01-NOV-1995 (Rel. 32, Last annotation update)
CC DE Hypothetical 40.1 kDa protein T20B12.4 in chromosome III.
CC GN T20B12.4.
CC OS Caenorhabditis elegans.
CC OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
CC OC Rhabditidae; Peloderinae; Caenorhabditis.
CC OX NCBI_TaxID=6239;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=Bristol N2;
CC RA Waterston R.;
CC RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
CC
CC -!- SIMILARITY: STRONG, TO C.ELEGANS K06H7.2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U10401; AAA19057.1; -
CC DR

```

DR WormPep: T20B12.4; CE01410.  
KW Hypothetical protein.  
SQ SEQUENCE 335 AA; 40139 MW; 9DE51B219062E8E7 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 335;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 DFSQIKN 483  
|||||  
Db 183 DFSQIKN 189

## RESULT 45

RUVB\_RICCN STANDARD; PRT; 342 AA.  
AC Q92187;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Holliday junction DNA helicase ruvb.  
GN RUVB OR RC0533.  
OS Rickettsia conorii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI\_taxid=781;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=Malish 7;  
RX MEDLINE=21442074; PubMed=11557893;

RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,  
RA Sanson D., Roux V., Cossart J., Weissbach J., Claverie J.-M.,  
RA Raoult D.;  
PT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";  
RL Science 293:2093-2098(2001).

CC -!- FUNCTION: The ruva-ruvb complex in the presence of ATP renatures  
CC cruciform structure in supercoiled DNA with palindromic sequence,  
CC indicating that it may promote strand exchange reactions in  
CC homologous recombination. RuvAB is an helicase that mediates the  
CC Holliday junction migration by localized denaturation and  
CC reannealing (By similarity).  
CC -!- SUBUNIT: Forms a complex with ruva (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE RUVB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AE008615; AAL03071.1; ALT\_INIT.  
DR InterPro; IPR003593; AAA\_ATPase.

DR InterPro; IPR003959; AAA\_ATPase\_cent.

DR InterPro; IPR004605; RuVB.

DR Pfam; PF00004; AAA; 1.

DR SMART; SM00382; AAA; 1.

DR TIGRFAMs; TIGR00635; ruvb; 1.

KW DNA repair; SOS response; ATP-binding; DNA recombination; Helicase;  
KW Complete proteome.

FT NP\_BIND 57 64 ATP (POTENTIAL).  
SQ SEQUENCE 342 AA; 38309 MW; 83891BD90C829C28 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 342;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 681 RASKLLD 687  
|||||  
Db 190 RASKLLD 196

## RESULT 46

RUVB\_RICPR STANDARD; PRT; 342 AA.  
AC Q92DE5;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Holliday junction DNA helicase ruvb.  
GN RUVB OR RP386.

OS Rickettsia prowazekii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI\_taxid=782;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=Madrid E;  
RX MEDLINE=99039499; PubMed=9823893;

RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,

RA Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,

RA Eriksson A.-S., Winkler H.H., Kurland C.G.;

RT "The genome sequence of Rickettsia prowazekii and the origin of

RT mitochondria";

RL Nature 396:133-140(1998).

CC -!- FUNCTION: THE RUVA-RUVB COMPLEX IN THE PRESENCE OF ATP RENATURES

CC CRUCIFORM STRUCTURE IN SUPERCOILED DNA WITH PALINDROMIC SEQUENCE,

CC INDICATING THAT IT MAY PROMOTE STRAND EXCHANGE REACTIONS IN

CC HOMOLOGOUS RECOMBINATION. RUVAB IS AN HELICASE THAT MEDIATES THE

CC HOLLIDAY JUNCTION MIGRATION BY LOCALIZED DENATURATION AND

CC REANNEALING (BY SIMILARITY).

CC -!- FUNCTION: PARTICIPATES TO UV-TOLERANCE OF SYNECHOCYSTIS PCC 6803.

CC -!- SUBUNIT: FORMS A COMPLEX WITH RUVA.

CC -!- SIMILARITY: BELONGS TO THE RUVB FAMILY.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AJ235271; CAA14843.1; --

DR InterPro; IPR003593; AAA\_ATPase.

DR InterPro; IPR003959; AAA\_ATPase\_cent.

DR Pfam; PF00004; AAA; 1.

DR SMART; SM00382; AAA; 1.

DR TIGRFAMs; TIGR00635; ruvb; 1.

KW DNA repair; SOS response; ATP-binding; DNA recombination; Helicase;

KW Complete proteome.

FT NP\_BIND 57 64 ATP (POTENTIAL).  
SQ SEQUENCE 342 AA; 38435 MW; D2F9290E84079BA5 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 342;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 681 RASKLLD 687  
|||||  
Db 190 RASKLLD 196

## RESULT 47

NO08\_PARDE STANDARD; PRT; 345 AA.  
AC P29920;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE NADH-ubiquinone oxidoreductase chain 8 (EC 1.6.5.3) (NADH

DE dehydrogenase 1, chain 8) (NDH-1, chain 8).

GN NO08.

OS Paracoccus denitrificans.

```

CC Bacteria: Proteobacteria; alpha subdivision; Rhodobacter group;
CC Paracoccus.
OX NCBI_TaxID=266;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 13543;
RX MEDLINE=93136200; PubMed=8422400;
RA Xu X., Matsuno-Yagi A., Yagi T.;
RT "DNA sequencing of the seven remaining structural genes of the gene
RT cluster encoding the energy-transducing NADH-quinone oxidoreductase
RT of Paracoccus denitrificans.";
RL Biochemistry 32:968-981(1993).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SUBUNIT: COMPOSED OF 14 DIFFERENT SUBUNITS, SUBUNITS N007-14
CC CONSTITUTE THE MEMBRANE SECTOR OF THE COMPLEX.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L02354; AAA25592.1; -
DR PIR: C45456;
DR InterPro: IPR001694; Resp_NADH_dhl.
DR Pfam: PF00146; NADHdh; 1.
DR PROSITE; PS00667; COMPLEX1_ND1_1; 1.
DR PROSITE; PS00668; COMPLEX1_ND1_2; 1.
KW Oxidoreductase; NAD; Ubiquinone; Transmembrane.
FT TRANSMEM 15 35 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 115 135 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 190 210 POTENTIAL.
FT TRANSMEM 253 273 POTENTIAL.
FT TRANSMEM 278 298 POTENTIAL.
FT TRANSMEM 309 329 POTENTIAL.
SQ SEQUENCE 345 AA; 38751 MW; E33B667E569506B4 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 345;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 671 PFLSML 677
DB 88 PFLSML 94
|||||

RESULT 48
YJ9L YEAST
ID YJ9L YEAST STANDARD; PRT; 358 AA.
AC P47175;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 41.2 kDa protein in RPS4A-BAR2 intergenic region.
GN YJ9L147W OR J7204.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Scarce T.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: BELONGS TO THE HSF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z49646; CAA89680.1; -
DR HSSP; P22121; ZHTS.
DR TRANSFAC; T03449; -
DR SGD; S0003908; YJ9L147W.
DR InterPro: IPR002323; HSF_DNA_bind.
DR InterPro: IPR002341; HSF_ETs.
DR Pfam; PF00447; HSF_DNA-bind; 1.
DR ProDom; PD001788; HSF_DNA_bind; 1.
DR SMART; SM00415; HSF; 1.
DR PROSITE; PS00434; HSF_DOMAIN; FALSE_NEG.
KW Hypothetical protein; Nuclear protein; DNA-binding.
FT DNA_BIND 12 116 BY SIMILARITY.
SQ SEQUENCE 358 AA; 41192 MW; E2964157FC952D42 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 358;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 QLITLLS 611
DB 347 QLITLLS 353
|||||

RESULT 49
BIOB_SYNY3
ID BIOB_SYNY3 STANDARD; PRT; 362 AA.
AC P73538;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Biotin synthase (EC 2.8.1.6) (Biotin synthetase).
GN BIOB OR SLR1364.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Watsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Tamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- CATALYTIC ACTIVITY: Dethiobiotin + sulfur = biotin.
CC -1- PATHWAY: Biotin biosynthesis; last step.
CC -1- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D90907; BAA17578.1; -
DR InterPro: IPR002684; Biotin_synth.
DR Pfam; PF01792; Biotin_synth; 1.
DR TIGFAMS; TIGR00433; bioB; 1.
KW Biotin biosynthesis; Iron-sulfur; Transferrase; Complete proteome.
FT METAL 88 88 IRON-SULFUR (POTENTIAL).

```



FT METAL 92 92 IRON-SULFUR (POTENTIAL).  
FT METAL 95 95 IRON-SULFUR (POTENTIAL).  
SQ SEQUENCE 362 AA; 33349 MW; BF7A16FB1CB68B4 CRC64;  
  
Query Match 0.6%; Score 7; DB 1; Length 362;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 994 KTEAEIL 1000  
Db 112 KTEAEIL 118  
  
RESULT 50  
YPC4\_CABEL STANDARD; PRT; 371 AA.  
AC Q1181;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 41.9 kDa protein C05D10.4 in chromosome III.  
GN C05D10.4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Du Z.  
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U13645; AAA20988.1; -  
DR WormPep; C05D10.4; CE01127.  
DR InterPro; IPR001660; SAM.  
DR Pfam; PF00536; SAM; 1  
DR SMART; SM00454; SAM; 1.  
DR PROSITE; PS0105; SAM\_DOMAIN; 1.  
KW Hypothetical protein.  
FT DOMAIN 223 286  
SQ SEQUENCE 371 AA; 41906 MW; 6CE79B87D8DD390C CRC64;

Query Match 0.6%; Score 7; DB 1; Length 371;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 115 GSTEVS 121  
Db 113 GSTEVS 119

Search completed: November 6, 2002, 03:35:05  
Job time : 100 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 16:47:42 ; Search time 702 Seconds  
(without alignments)  
11968.954 Million cell updates/sec

Title: US-09-782-874-1  
Perfect score: 3731  
Sequence: 1 GAATATCTTACTACTT.....AGTTTCATCTTCTCTAA 3731

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3731	100.0	3731	22 AAD04370	Tomato RNA-directed
2	3731	100.0	3731	23 AAS17837	Tomato RNA-directed
3	3721.4	99.7	3731	23 AAS17845	Tomato RNA-directed
4	1396	37.4	3807	21 AAC63742	Soybean RNA-directed
5	987.8	26.5	3737	21 AAC63739	Maize RNA-directed
6	345	9.2	3901	21 AAC63740	Maize RNA-directed
7	305.8	8.2	3591	22 AAH77704	Nucleotide sequenc
8	304.2	8.2	2816	21 AAC63741	Rice RNA-directed
9	282	7.6	1281	21 AAC63738	Maize RNA-directed

10	276.2	7.4	740	21	AAC63743	Maize RNA-directed
11	220.4	5.9	6863	22	AAH77703	Nucleotide sequenc
12	106.8	2.9	282	24	ABL71773	Corn tassell-derive
13	95.4	2.6	274	24	ABL71722	Corn tassell-derive
14	85.8	2.3	349	21	AAC63746	Soybean RNA-direct
15	71.4	1.9	510	21	AAC63745	Soybean RNA-direct
c 16	57.2	1.5	906	17	AAT18078	Potato calmodulin
17	49	1.3	8045	21	AAAG5171	Neurospora crassa
18	41.2	1.1	339	21	AAC18542	Human secreted pro
c 19	41.2	1.1	7156	24	ABL59541	Human interleukin
c 20	40.4	1.1	23071	22	AAS26699	Human genomic DNA
21	40.2	1.1	1904	22	AAK05148	Human brain expres
22	40.2	1.1	14041	22	AAH48024	Internal control B
c 23	40	1.1	4590	22	AAH48065	yeast AOD9604-asso
24	40	1.1	162450	21	AZ86967	Retinoblastoma bin
25	39.4	1.1	10716	24	ABL33418	Human immune syste
26	39.2	1.1	2073	23	AAS54725	Staphylococcus aur
27	39.2	1.1	2109	23	AAS51973	Staphylococcus aur
28	39.2	1.1	6012	24	ABL70327	Chemically treated
29	39.2	1.1	6012	24	AS61274	Human gene regulat
30	39.2	1.1	6012	24	ABK31370	Signal transductio
31	39	1.0	7560	24	ABL33223	Human immune syste
32	38.8	1.0	11790	24	ABL32542	Human immune syste
33	38.6	1.0	2570	23	ABL22608	Drosophila melanog
34	38.6	1.0	4803	23	ABL16961	Drosophila melanog
35	38.6	1.0	6079	24	ABL32421	Human immune syste
36	38.6	1.0	6225	23	ABL20912	Drosophila melanog
37	38.6	1.0	7027	23	ABL16960	Drosophila melanog
38	38.4	1.0	5179	24	ABL33996	Human immune syste
39	38.4	1.0	5179	24	ABK28403	DNA transcription
40	38.2	1.0	1537	21	AAC39240	Arabidopsis thalia
41	38.2	1.0	15667	24	ABL34146	Human immune syste
42	38.2	1.0	113515	24	ABL34174	Human immune syste
c 43	38	1.0	540	24	ABN64825	Human cancer relat
c 44	38	1.0	4236	24	AAI67659	Nucleotide sequenc
c 45	38	1.0	5191	21	AAH20989	Human low adenosin

## ALIGNMENTS

RESULT 1  
AAD04370

ID AAD04370 standard; cdna; 3731 BP.

AC AAD04370;

XX AAD04370;

XX 04-JUL-2001 (first entry)

XX Tomato RNA-directed RNA polymerase (RdRP) cdna.

DE DE

KW Tomato: gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;

KW transgenic plant; tissue culture; plant breeding; therapy; C-protein; ss.

XX Lycopersicon esculentum.

OS OS

XX Key Location/Qualifiers

XX CDS 194..3538

XX /\*tag= a

XX /product= "Tomato C-protein having RNA-directed RNA

XX polymerase (RdRP) activity"

XX US6218142-B1.

XX 17-APR-2001.

XX 05-MAR-1997; 97US-0811583.

XX 05-MAR-1997; 97US-0811583.

XX (WASS/) WASENEGGER M.

XX (RIED/) RIEDEL L.

PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;  
XX  
DR WPI; 2001-289830/30.  
DR P-PSDB; AAE00897.

aa New nucleic acid molecules encoding polypeptides with RNA-directed RNA  
PT polymerase enzymatic activity, useful in modulating gene expression in  
PT plants, humans and animals, as well as in plant cell/tissue cultures or  
PT plant breeding -

PS Claim 1; Column 25-34; 31pp; English.

The present sequence is a cDNA encoding tomato C-protein having RNA-directed RNA polymerase (RdRP) activity. This protein is capable of RNA-directed RNA synthesis, thus using RNA as a template for synthesizing complementary RNA molecules. RdRP nucleic acid is useful for modulating gene expression in plants, humans and animals. This may lead to various physiological, developmental and/or morphological changes. Transgenic plants containing RdRP nucleic acid is especially useful in plant cell or tissue cultures and in plant breeding. RdRP is useful in gene therapy, particularly for treating a disease that is caused by the undesirable expression or overexpression of a gene.

Sequence 3731 BP: 1064 A; 669 C; 849 G; 1149 T; 0 other;

```

Query Match      100.0%; Score 3731; DB 22; Length 3731;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy	1	GAAATATTCTTTACTTACCTTCACCAGGATTGACATCATCTCCCTCAAGTCCTTTGTGT	60
Db	1	GAAATATTCTTTACTTACCTTCACCAGGATTGACATCATCTCCCTCAAGTCCTTTGTGT	60
Qy	61	GTGTGATAATAAATTTGGTTGTGCTCAGTTTTCAGTTTCAGTGCACACTGCTGGTAGTTTTAT	120
Db	61	GTGTGATAATAAATTTGGTTGTGCTCAGTTTTCAGTTTCAGTGCACACTGCTGGTAGTTTTAT	120
Qy	121	TTTTCGATTAACCTTCAGGGGTATTCAGTTGTGTTAGCATTTGAAGTCGAACTGCACACTT	180
Db	121	TTTTCGATTAACCTTCAGGGGTATTCAGTTGTGTTAGCATTTGAAGTCGAACTGCACACTT	180
Qy	181	GGAAATTTGGCTACATCGGAAAGACAATTCAGTTTTCGGATTCCCTTATCTTCTCTCTGC	240
Db	181	GGAAATTTGGCTACATCGGAAAGACAATTCAGTTTTCGGATTCCCTTATCTTCTCTCTGC	240
Qy	241	GGAAGTGGTTTAAGTCATCTTATAGAGAAATATACAGGATATGGAAGTGTATGTGCAATGGA	300
Db	241	GGAAGTGGTTTAAGTCATCTTATAGAGAAATATACAGGATATGGAAGTGTATGTGCAATGGA	300
Qy	301	GGTTTAACAGTCCAAAGGAGGATCTAGAGCATTTGCCAAAAGTTCAAATTTGCCGACACAT	360
Db	301	GGTTTAACAGTCCAAAGGAGGATCTAGAGCATTTGCCAAAAGTTCAAATTTGCCGACACAT	360
Qy	361	AAGTGTCTGACAAAATCATCACTTTGGCTTAATACAGGCTGTATTTTGGCTCTCTCTTATTT	420
Db	361	AAGTGTCTGACAAAATCATCACTTTGGCTTAATACAGGCTGTATTTTGGCTCTCTCTTATTT	420
Qy	421	GAAGGCTTTGGAAATGAAAACATGATATTCTCCAAGTCGGGATATATGGATTCAGATGGA	480
Db	421	GAAGGCTTTGGAAATGAAAACATGATATTCTCCAAGTCGGGATATATGGATTCAGATGGA	480
Qy	481	TGGCATAACTTTGAATTTCCGATGTGCAGATATCAGATGACAAAGTTTGCAGTGTGGGAAG	540
Db	481	TGGCATAACTTTGAATTTCCGATGTGCAGATATCAGATGACAAAGTTTGCAGTGTGGGAAG	540
Qy	541	TACAGAAGTTTCAATTCAAATTTGGCATTCGATTGAAGAAATTTTTTCTCTTTTATCTAG	600
Db	541	TACAGAAGTTTCAATTCAAATTTGGCATTCGATTGAAGAAATTTTTTCTCTTTTATCTAG	600
Qy	601	TGSTTCAGCTGACTATAAACCTTCAGGTTTCAATATGAAAAATATATGCGAGTTGTGCTCCA	660
Db	601	TGSTTCAGCTGACTATAAACCTTCAGGTTTCAATATGAAAAATATATGCGAGTTGTGCTCCA	660

|||||  
Db 1741 GGTTCATGGAACCAAGCTATGCTTTTCTGATGGAAATGGTAAATATCTGGTGACTTTGC 1800  
QY 1801 TCATAGAGTTGCCCTCAAAATGTGGCTTCAATATACCCCATCTGCTTTCCAGATTCGTTA 1860  
Db 1801 TCATAGAGTTGCCCTCAAAATGTGGCTTCAATATACCCCATCTGCTTTCCAGATTCGTTA 1860  
QY 1861 TGGTGGATATAAAGGTGTGGTGTGGTGTGGATCCGGATTCATCAATGAAGTTGCTTTGAG 1920  
Db 1861 TGGTGGATATAAAGGTGTGGTGTGGTGTGGATCCGGATTCATCAATGAAGTTGCTTTGAG 1920  
QY 1921 AAAGAGCATGTCGAATATGAATCAGACAAACATAAAGTTAGATGTCCTTGGATGGAGCAA 1980  
Db 1921 AAAGAGCATGTCGAATATGAATCAGACAAACATAAAGTTAGATGTCCTTGGATGGAGCAA 1980  
QY 1981 ATATCAGCCTGTTTATCTTAATCGTCAACTGATTAAGCTCTTGTCTACACTTGGAGTGA 2040  
Db 1981 ATATCAGCCTGTTTATCTTAATCGTCAACTGATTAAGCTCTTGTCTACACTTGGAGTGA 2040  
QY 2041 AGATGAAGTTCGGAACAGAGCAAAAGGAAGCTGTAGATCAGCTTGTATCTTCTTTCGA 2100  
Db 2041 AGATGAAGTTCGGAACAGAGCAAAAGGAAGCTGTAGATCAGCTTGTATCTTCTTTCGA 2100  
QY 2101 TGATCTTTGAAGGCACAGGAGGCTTTGGAATGTAGTCTCTCTGGAGAGAACACTAATAT 2160  
Db 2101 TGATCTTTGAAGGCACAGGAGGCTTTGGAATGTAGTCTCTCTGGAGAGAACACTAATAT 2160  
QY 2161 TCTCAAGCAATGCTAAACTGTGGTTATAAGCTGTATGAGCCCTTCTTTCAATGAT 2220  
Db 2161 TCTCAAGCAATGCTAAACTGTGGTTATAAGCTGTATGAGCCCTTCTTTCAATGAT 2220  
QY 2221 GTTGCAAACTTCCGCGCATCCAAAGTTGCTCGATTGGGACTAGATCAAGAATATTTAT 2280  
Db 2221 GTTGCAAACTTCCGCGCATCCAAAGTTGCTCGATTGGGACTAGATCAAGAATATTTAT 2280  
QY 2281 TCCAAATGAAGAACATGATGGATGTTGGATGAATCCAGAACTTGGATATGGTCA 2340  
Db 2281 TCCAAATGAAGAACATGATGGATGTTGGATGAATCCAGAACTTGGATATGGTCA 2340  
QY 2341 GGTGTTGTTGCTAGTTTACTGTGCTGGACATGGAGAGTTTCTGACGATTTACATCCATT 2400  
Db 2341 GGTGTTGTTGCTAGTTTACTGTGCTGGACATGGAGAGTTTCTGACGATTTACATCCATT 2400  
QY 2401 TAATACAGCAGATCCACCAACAGTAATTTCTGAGGGAATGTTGTTGTCGAAA 2460  
Db 2401 TAATACAGCAGATCCACCAACAGTAATTTCTGAGGGAATGTTGTTGTTGTCGAAA 2460  
QY 2461 AAATCCATGCTTGCATCCTGGTGATATCGTGTTTTAAAGGCTGTAATGTTTCGAGCGCT 2520  
Db 2461 AAATCCATGCTTGCATCCTGGTGATATCGTGTTTTAAAGGCTGTAATGTTTCGAGCGCT 2520  
QY 2521 GCACCACTGGTAGATGTTGTTGATTCCTCAGAAAGGAAAAGACCTCATCCGAATGA 2580  
Db 2521 GCACCACTGGTAGATGTTGTTGATTCCTCAGAAAGGAAAAGACCTCATCCGAATGA 2580  
QY 2581 ATGTTCTGGAGTGATTTGGATGGGATATCTACTTTGTTGCTGGATCAAGACATGAT 2640  
Db 2581 ATGTTCTGGAGTGATTTGGATGGGATATCTACTTTGTTGCTGGATCAAGACATGAT 2640  
QY 2641 CCGGCCAAGGCAAGTCCAGCGGATGGAATATCTCCAGCACCAGCATACAGTTGGACCA 2700  
Db 2641 CCGGCCAAGGCAAGTCCAGCGGATGGAATATCTCCAGCACCAGCATACAGTTGGACCA 2700  
QY 2701 TGATGTCACAATTTGAGGAAGTTCAAGAGTACTTCCCACTATTTGTTGAATGACAGTTT 2760  
Db 2701 TGATGTCACAATTTGAGGAAGTTCAAGAGTACTTCCCACTATTTGTTGAATGACAGTTT 2760  
QY 2761 GGGAAATCATAGCAAAATGCCCATGCTCGTATTGTCAGACAGAGAACCTGATATGGCCATGAG 2820  
Db 2761 GGGAAATCATAGCAAAATGCCCATGCTCGTATTGTCAGACAGAGAACCTGATATGGCCATGAG 2820  
QY 2821 TGATCCATGCAAAAACCTTCTGAGCTCTTTTCAATTGAGTGGACTTTCCAAAGACTGG 2880  
|||||

Db 2821 TGATCCATGCAAAAACCTTCTGAGCTCTTTTCAATTGAGTGGACTTTTCCAAAGACTGG 2880  
QY 2881 TGTTCGCCGTGAATATACCATCTCAGTTGGCCCTTAAGATATACCAGACTTTCATGGATA 2940  
Db 2881 TGTTCGCCGTGAATATACCATCTCAGTTGGCCCTTAAGATATACCAGACTTTCATGGATA 2940  
QY 2941 GCCGGCAACAGCAGCAGTATATCTCAGAAAGAGTTATTGGAAGCTTTTCCAGGAAAGTGA 3000  
Db 2941 GCCGGCAACAGCAGCAGTATATCTCAGAAAGAGTTATTGGAAGCTTTTCCAGGAAAGTGA 3000  
QY 3001 GGCAAAAGCACCCTCAGGCTAGCTCTATCGGACCTTCCACAAGAGATGTTCCAAAGGAGATC 3060  
Db 3001 GGCAAAAGCACCCTCAGGCTAGCTCTATCGGACCTTCCACAAGAGATGTTCCAAAGGAGATC 3060  
QY 3061 ATATGATGCTGATATGAAGTTGATGGAATTTGAAGATTACATTGACGAAGCTTTTGACTA 3120  
Db 3061 ATATGATGCTGATATGAAGTTGATGGAATTTGAAGATTACATTGACGAAGCTTTTGACTA 3120  
QY 3121 CAAACTGAATATGACAAACAGCTGGTAAATTTAATGGACTACTATGGCATATAAAGACAGA 3180  
Db 3121 CAAACTGAATATGACAAACAGCTGGTAAATTTAATGGACTACTATGGCATATAAAGACAGA 3180  
QY 3181 GGCTGAATATCTTGTAGTGGTGGCATTTAAGAGGCATCAAAAACCTTTTACCGCAGAAAAGA 3240  
Db 3181 GGCTGAATATCTTGTAGTGGTGGCATTTAAGAGGCATCAAAAACCTTTTACCGCAGAAAAGA 3240  
QY 3241 TGCTGAGGCCATTAATGCTGTGAGGGCTTTGAGGAAGAGGCAAGAGCCTGGTTCAA 3300  
Db 3241 TGCTGAGGCCATTAATGCTGTGAGGGCTTTGAGGAAGAGGCAAGAGCCTGGTTCAA 3300  
QY 3301 GAGCGCTAATGATATAGATGACATGTTACCAAGGCTTGGCTTGGTACCACGTTACATA 3360  
Db 3301 GAGCGCTAATGATATAGATGACATGTTACCAAGGCTTGGCTTGGTACCACGTTACATA 3360  
QY 3361 TCATCTACATATTGGGGTGGCTACAATCAGGGGTTGAAAAGAGCTCATTTTCATTAGCTT 3420  
Db 3361 TCATCTACATATTGGGGTGGCTACAATCAGGGGTTGAAAAGAGCTCATTTTCATTAGCTT 3420  
QY 3421 TCCCTGGTGTGTTTATGACCAAGCTTAATCCAGATTAAAGAGCAAAAGCAGTAAACAGGCC 3480  
Db 3421 TCCCTGGTGTGTTTATGACCAAGCTTAATCCAGATTAAAGAGCAAAAGCAGTAAACAGGCC 3480  
QY 3481 AGTTCTCAACTGTTCATCTCAGGGCTCAACTGAGTCACAGATTAGTTGTTGAAATGAGA 3540  
Db 3481 AGTTCTCAACTGTTCATCTCAGGGCTCAACTGAGTCACAGATTAGTTGTTGAAATGAGA 3540  
QY 3541 TTCAGTCGAGCGTTAAGCTGATATATATATAGTAAATAGGTTGATCATATAAGAAAAC 3600  
Db 3541 TTCAGTCGAGCGTTAAGCTGATATATATATAGTAAATAGGTTGATCATATAAGAAAAC 3600  
QY 3601 TGTATGCTATGTTGACTACCTTTTGTCTTTTAAACTGCATGAAGCTGCAACATATATGC 3660  
Db 3601 TGTATGCTATGTTGACTACCTTTTGTCTTTTAAACTGCATGAAGCTGCAACATATATGC 3660  
QY 3661 AGTACTCTAAGAAACAGATGATGACAGTAAAGTAAATATGATGATTTGATGTTGATTCATC 3720  
Db 3661 AGTACTCTAAGAAACAGATGATGACAGTAAAGTAAATATGATGATTTGATGTTGATTCATC 3720  
QY 3721 TTTCTTCTAAA 3731  
Db 3721 TTTCTTCTAAA 3731  
RESULT 2  
AAS17837  
ID AAS17837 standard; cDNA; 3731 BP.  
XX AAS17837;  
AC AAS17837;  
XX 08-MAY-2002 (first entry)  
XX Tomato RNA-directed RNA polymerase (RdRP) cDNA.  
XX

KW Tomato: RdRP: RNA-directed RNA polymerase; in vitro transcription;  
 KW cytosolic; viricide; RNA synthesis inhibitor; antibody; immunogen;  
 KW transgenic plant; transgenic animal; cancer; viral infection;  
 KW immunoprecipitation; immunolocalisation; ss; gene therapy.

XX Lycopersicon esculentum.

Key Location/Qualifiers  
 CDS 194..3538  
 /\*tag= a  
 /product= "RdRP protein"  
 /EC\_number= "2.7.7.48"

US2001023067-A1.

20-SEP-2001.

08-FEB-2001; 2001US-0782874.

05-MAR-1997; 97US-0811583.

(WASS/) WASSNEGGER M.  
 (RIED/) RIEDEL L.

Wassenegger M, Riedel L, Schiebel W, Sanger HL;

WPI; 2001-595798/67.

P-PSDB: AAU10006.

New nucleic acid molecule encoding a polypeptide having the enzymatic activity of RNA-directed RNA polymerase, for modulating gene expression and treating cancer and virus infection in human and animals

Claim 1; Page 15-19; 34pp; English.

This sequence represents a cDNA encoding the tomato RNA-directed RNA polymerase (RdRP) protein of the invention. The invention comprises the nucleic acid and protein sequences of RdRP. The protein of the invention can catalyze in vitro transcription of short single stranded RNAs into DNA molecules, this transcription can be either primed by RNA or DNA oligonucleotides or be unprimed. The protein may have cytosolic or virucide activities. The sequences of the invention may be used in gene therapy or as an RNA directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template nucleic acid molecule derived from a nucleic acid molecule which causes a disease are useful for treating a disease caused by the undesired expression or overexpression of a nucleic acid molecule in a human, rat or mouse, by administering the molecules. This system can be used in the preparation of a pharmaceutical composition and for inhibiting expression of any desired gene by transferring the RdRP system to organisms that either lack a comparable mechanism or do not sufficiently express their own RdRP. An antibody or an antagonist or inhibitor to the protein are useful for inhibiting RNA directed RNA synthesis and for ensuring stable heterologous, gene expression in transgenic organisms. The sequence is useful for probes and/or for the control of gene expression, as primers for amplification of nucleic acid molecules and as tools for the detection of expression of the cDNA molecules. Additionally, nucleotide and protein sequences are useful for suppression of undesired gene expression in humans and animals. The RdRP is useful as a therapeutic agent for the control of cancer and virus infection in humans and animals and the antibody is useful for immunoprecipitation or immunolocalisation of the protein, identification of polypeptides interacting with it and screening expression libraries.

Sequence 3731 BP; 1064 A; 569 C; 849 G; 1149 T; 0 other;

Query Match 100.0%; Score 3731; DB 23; Length 3731;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAATATCTTTACTTACACAGGATGACATCATCTCCCTCAAGTCITTTGT 60  
 |||||  
 DB 1 GAAATATCTTTACTTACACAGGATGACATCATCTCCCTCAAGTCITTTGT 60

QY 61 GTTGTGATATAAATTTGGTTGGTTTTCAGTTTTCAGTCTACTGCTGGTAGTTTAT 120  
 DB 61 GTTGTGATATAAATTTGGTTGGTTTTCAGTTTTCAGTCTACTGCTGGTAGTTTAT 120  
 QY 121 TTTGCATAACTTCAGGGGGTATTCCAGTTGGTGTAGCATTTGAAAGTGAACACTGCAC 180  
 DB 121 TTTGCATAACTTCAGGGGGTATTCCAGTTGGTGTAGCATTTGAAAGTGAACACTGCAC 180  
 QY 181 GGAATTTGGCTACATGGGAAAGACAATTGAGTTTTCGGATTCCTTATCTCTCTCG 240  
 DB 181 GGAATTTGGCTACATGGGAAAGACAATTGAGTTTTCGGATTCCTTATCTCTCTCG 240  
 QY 241 GGAAGTGGTTAAGTCAATTTAGAGAAATATACAGATATGGAACCTGTATGTGCATTGA 300  
 DB 241 GGAAGTGGTTAAGTCAATTTAGAGAAATATACAGATATGGAACCTGTATGTGCATTGA 300  
 QY 301 GGTAAACAGTCCAAAGGAGGATCTAGAGCAATTTGCCAAGTTCATTTGCCGACACAT 360  
 DB 301 GGTAAACAGTCCAAAGGAGGATCTAGAGCAATTTGCCAAGTTCATTTGCCGACACAT 360  
 QY 361 AAGTGCTGACAAATCATCATCTTTGGCTAATAACAGCTGTATTTGGGCTCTCTCTATT 420  
 DB 361 AAGTGCTGACAAATCATCATCTTTGGCTAATAACAGCTGTATTTGGGCTCTCTCTATT 420  
 QY 421 GAAGGCTTGGGAAATGAAAACACTGATATTTGCCAACTCCGGGCATATGTGCATGGA 480  
 DB 421 GAAGGCTTGGGAAATGAAAACACTGATATTTGCCAACTCCGGGCATATGTGCATGGA 480  
 QY 481 TGGCATTAACATTTGAAATTTGGGATGTCAGATATGAGATGACAAAGTTTGGGAGAG 540  
 DB 481 TGGCATTAACATTTGAAATTTGGGATGTCAGATATGAGATGACAAAGTTTGGGAGAG 540  
 QY 541 TAGCAGAAGTTTCAATTTCAATTTGGCATTTGATTGAAGAAATTTTTTCTTTTATCTAG 600  
 DB 541 TAGCAGAAGTTTCAATTTCAATTTGGCATTTGATTGAAGAAATTTTTTCTTTTATCTAG 600  
 QY 601 TGGTTGAGTCACTATAAATTTGAGTTTTCAGTTTTCATATGAAATATATGACAGTTGTC 660  
 DB 601 TGGTTGAGTCACTATAAATTTGAGTTTTCAGTTTTCATATGAAATATATGACAGTTGTC 660  
 QY 661 TCGTCCATATGGTCAAAATGCTCAGTTTCCCTCATACAGTTATTTGGTGGCTCCTCGGAT 720  
 DB 661 TCGTCCATATGGTCAAAATGCTCAGTTTCCCTCATACAGTTATTTGGTGGCTCCTCGGAT 720  
 QY 721 CTATAAGAGACTTGAAGAACTCTGTTATAGTTTCTTAAAGAACTCCTCATGATCAGTG 780  
 DB 721 CTATAAGAGACTTGAAGAACTCTGTTATAGTTTCTTAAAGAACTCCTCATGATCAGTG 780  
 QY 781 GGTGAGGACACAGATTTCCCTCATCTTGGATAGGCTATCTTCTAGCTTATGTTTGA 840  
 DB 781 GGTGAGGACACAGATTTCCCTCATCTTGGATAGGCTATCTTCTAGCTTATGTTTGA 840  
 QY 841 GTTCCGTAGGGGTTCGTCTTCCAAATTTTCGAGGAAAGTTTTTCCACTATCGAAGC 900  
 DB 841 GTTCCGTAGGGGTTCGTCTTCCAAATTTTCGAGGAAAGTTTTTCCACTATCGAAGC 900  
 QY 901 TGAACAATATTAATTTACAGACTGTTTCACTTTTTCGTTCTCTCAAAATTCGGCTCT 960  
 DB 901 TGAACAATATTAATTTACAGACTGTTTCACTTTTTCGTTCTCTCAAAATTCGGCTCT 960  
 QY 961 GGTTCCTCAATGTCAGCTCCGGAAGAAATTTCAATTCCTACAGATTTGTTGTTAAAT 1020  
 DB 961 GGTTCCTCAATGTCAGCTCCGGAAGAAATTTCAATTCCTACAGATTTGTTGTTAAAT 1020  
 QY 1021 TAGTTCTTTGGTACAGCATGGATGATACCTGGCCGAGCAATTAATGCTACTTTTCCG 1080  
 DB 1021 TAGTTCTTTGGTACAGCATGGATGATACCTGGCCGAGCAATTAATGCTACTTTTCCG 1080  
 QY 1081 ATTAGTTGATCTCGAAGAGAAATGCGCATGCAATGAGCATGCCCTTAGAGAACTGTA 1140  
 DB 1081 ATTAGTTGATCTCGAAGAGAAATGCGCATGCAATGAGCATGCCCTTAGAGAACTGTA 1140  
 QY 1141 CTATATAAGAGAGTGTGTTATGATCCCGTAGGCTCCTCAGTGAGCAGTATGATGGTA 1200

Db	1141	CTATATAAAGGAGTGTCTGTATATGATCCCGTGGTGGCTCACTGAGCAGTATCATGGTA	1200
Qy	1201	TCFCAAGGTTAGACAACCTCCAAAATCTCCGTCCATCACTTTAGATCATGGTGTGGTGA	1260
Db	1201	TCFCAAGGTTAGACAACCTCCAAAATCTCCGTCCATCACTTTAGATCATGGTGTGGTGA	1260
Qy	1261	TGPAAGAAGGTCCTAGTAACACCATGCAAAAGTTATTTTTGTGGTCCGAGAGTTAATGT	1320
Db	1261	TGPAAGAAGGTCCTAGTAACACCATGCAAAAGTTATTTTTGTGGTCCGAGAGTTAATGT	1320
Qy	1321	TTCCAATCGSGTTCCTCCGAATTTATCTGAAGACATAGATAACTTTCTTCGTGTCTCTTT	1380
Db	1321	TTCCAATCGSGTTCCTCCGAATTTATCTGAAGACATAGATAACTTTCTTCGTGTCTCTTT	1380
Qy	1381	TGTTGATGAGGAGTGGGAGAACTGTATCTTACAGACTTATACCAAAACCAAGTACTGG	1440
Db	1381	TGTTGATGAGGAGTGGGAGAACTGTATCTTACAGACTTATACCAAAACCAAGTACTGG	1440
Qy	1441	AAGTGGTGTCAAGACAACATCTATGAGAGGATCTTATCAACTCTCGGAAAGCCTTTGT	1500
Db	1441	AAGTGGTGTCAAGACAACATCTATGAGAGGATCTTATCAACTCTCGGAAAGCCTTTGT	1500
Qy	1501	AATTGTTGATAAAAAATTTGAAATTTCTTGCAATTTTCATFCGAGCCAGTTGCGGATAATTC	1560
Db	1501	AATTGTTGATAAAAAATTTGAAATTTCTTGCAATTTTCATFCGAGCCAGTTGCGGATAATTC	1560
Qy	1561	AGTGTGGATGTTGCATCAAGACCTTGGCCTTACTGCAAAATGATATAGAGCTTGGATGGG	1620
Db	1561	AGTGTGGATGTTGCATCAAGACCTTGGCCTTACTGCAAAATGATATAGAGCTTGGATGGG	1620
Qy	1621	TGATTTTTCGCAGATCAAGAATGTGCGAAAATATGCTGCCAGACTTGGTCAATCTTTTGG	1680
Db	1621	TGATTTTTCGCAGATCAAGAATGTGCGAAAATATGCTGCCAGACTTGGTCAATCTTTTGG	1680
Qy	1681	TTCTCCAGAGAGACTTTGAGTGTCTTAGCGATGAGATTCAGAGTTATCCCGATGTAAA	1740
Db	1681	TTCTCCAGAGAGACTTTGAGTGTCTTAGCGATGAGATTCAGAGTTATCCCGATGTAAA	1740
Qy	1741	GGTTCATGGAACCAAGCTATGTCTTCTGATGGAATTCGTAAAAATATCTGGTGACTTTGC	1800
Db	1741	GGTTCATGGAACCAAGCTATGTCTTCTGATGGAATTCGTAAAAATATCTGGTGACTTTGC	1800
Qy	1801	TCATAGAGTTGCCTCAAAATGTGGCCTTCAATATPACCCCATCTGCTTCCAGATTCGTTA	1860
Db	1801	TCATAGAGTTGCCTCAAAATGTGGCCTTCAATATPACCCCATCTGCTTCCAGATTCGTTA	1860
Qy	1861	TGTTGATATAAAGTGTGTGGTGTGATCCGGAATTCATCAATGAAGTTGTCTTTGAG	1920
Db	1861	TGTTGATATAAAGTGTGTGGTGTGATCCGGAATTCATCAATGAAGTTGTCTTTGAG	1920
Qy	1921	AAAGAGCATGCGAAATATGAATCAGACAACATAAAGTTAGATGTCCTGGATGGAGCAA	1980
Db	1921	AAAGAGCATGCGAAATATGAATCAGACAACATAAAGTTAGATGTCCTGGATGGAGCAA	1980
Qy	1981	ATATCAGCCTTGTTATCTTAATCGTCAACTGATTACGCTCTTGCTACACTTGGAGTAA	2040
Db	1981	ATATCAGCCTTGTTATCTTAATCGTCAACTGATTACGCTCTTGCTACACTTGGAGTAA	2040
Qy	2041	AGATGAAGTTCTCGAACAGAACAAAGAAAGCTGTAGATCAGCTTGATGTCTATCTTGCA	2100
Db	2041	AGATGAAGTTCTCGAACAGAACAAAGAAAGCTGTAGATCAGCTTGATGTCTATCTTGCA	2100
Qy	2101	TGATCTTTTGAAGGCACAGAGGCTTTGGAATTGATGTCTCTCGAGAGAACACTATAT	2160
Db	2101	TGATCTTTTGAAGGCACAGAGGCTTTGGAATTGATGTCTCTCGAGAGAACACTATAT	2160
Qy	2161	TCCTAAGGCAATGCTAAACTGTGGTTATTAAGCCTGTGATGCTGAGCCCTTTCTTCAATGAT	2220
Db	2161	TCCTAAGGCAATGCTAAACTGTGGTTATTAAGCCTGTGATGCTGAGCCCTTTCTTCAATGAT	2220
Qy	2221	GTTCCAAACCTTCCTGGGCAATCCAAAGTTGCTTCGATTTGGGACTAGATCAAGAATATTAT	2280

Db	2221	GTGCAAAACCTTCGGCGCATCCAAAGTTGCTGCATTTGGGACTAGATCAAGAATATTAT	2228
Qy	2281	TCCAAATGGAAGAACAAATGATGGATGTTTGGATGAATCCAGAACCCTTGGAAATATGTC	2340
Db	2281	TCCAAATGGAAGAACAAATGATGGATGTTTGGATGAATCCAGAACCCTTGGAAATATGTC	2340
Qy	2341	GGTGTGTTGTCAGTTTACTGGTGGTCTGGACATGGAGAGTTTCTGACGATTTACATCCATT	2400
Db	2341	GGTGTGTTGTCAGTTTACTGGTGGTCTGGACATGGAGAGTTTCTGACGATTTACATCCATT	2400
Qy	2401	TAATAACAGCAGATCCACCAACAGTAATTTCAATCTGAAGGGAATGTGTTGTGCAAA	2460
Db	2401	TAATAACAGCAGATCCACCAACAGTAATTTCAATCTGAAGGGAATGTGTTGTGCAAA	2460
Qy	2461	AAATCCATGCTGCATCCTGGTGGATATTCGTGTTTTAAAGCTGTAAATGTTCAGCGCT	2520
Db	2461	AAATCCATGCTGCATCCTGGTGGATATTCGTGTTTTAAAGCTGTAAATGTTCAGCGCT	2520
Qy	2521	GCACCATGGTAGATTGTGTTATTCCTCAGAAAGGAAAAAGACCTCATCCGGAATGA	2580
Db	2521	GCACCATGGTAGATTGTGTTATTCCTCAGAAAGGAAAAAGACCTCATCCGGAATGA	2580
Qy	2581	ATGTTCTGGGAGTGATTTGGATGGGATATCTACTGTTTCTGGGATCAAGACATGAT	2640
Db	2581	ATGTTCTGGGAGTGATTTGGATGGGATATCTACTGTTTCTGGGATCAAGACATGAT	2640
Qy	2641	CCGCCCAAGCAGTCCAGCCGATGGAATATCCTCCAGCACCCAGCATACAGTTTGACCA	2700
Db	2641	CCGCCCAAGCAGTCCAGCCGATGGAATATCCTCCAGCACCCAGCATACAGTTTGACCA	2700
Qy	2701	TGATGTCACAAATTTGAGGAAGTTGAAGAGTACTTACCACATATATTTGTGAATGACAGTTT	2760
Db	2701	TGATGTCACAAATTTGAGGAAGTTGAAGAGTACTTACCACATATATTTGTGAATGACAGTTT	2760
Qy	2761	GGGAATCATAGCAAAATGCCCATGCTGATTTTCGACAGAGAACTGATATGGCCATGAG	2820
Db	2761	GGGAATCATAGCAAAATGCCCATGCTGATTTTCGACAGAGAACTGATATGGCCATGAG	2820
Qy	2821	TGATCCATGCAAAAACCTTGCTGAGCTCTTTTCAATTCGAGTGGACTTTCCAAAGACTGG	2880
Db	2821	TGATCCATGCAAAAACCTTGCTGAGCTCTTTTCAATTCGAGTGGACTTTCCAAAGACTGG	2880
Qy	2881	TGTTCCCGCTGAATACCATCTCAGTTGGCCCTTAAAGAAATACCAGACTTTCATGGATAA	2940
Db	2881	TGTTCCCGCTGAATACCATCTCAGTTGGCCCTTAAAGAAATACCAGACTTTCATGGATAA	2940
Qy	2941	CCCGGCAAGACACCACTATATCTCAGAAAGATTTTGAAGAGCTTTTCAGGAAAGTGA	3000
Db	2941	CCCGGCAAGACACCACTATATCTCAGAAAGATTTTGAAGAGCTTTTCAGGAAAGTGA	3000
Qy	3001	GGACAAGCACCCTCAGCTAGCTCTATCCGACCTTCAAGAGATGTTTGAAGGAGATC	3060
Db	3001	GGACAAGCACCCTCAGCTAGCTCTATCCGACCTTCAAGAGATGTTTGAAGGAGATC	3060
Qy	3061	ATATGATGCTGATATGGAAGTTGATGGATTGAAGATTACATTCAGAAAGCTTTTGACTA	3120
Db	3061	ATATGATGCTGATATGGAAGTTGATGGATTGAAGATTACATTCAGAAAGCTTTTGACTA	3120
Qy	3121	CAAACTGAATATGACAACAAAGCTGGTAAATTTAATGGACTACTATGGCATTAACACAGA	3180
Db	3121	CAAACTGAATATGACAACAAAGCTGGTAAATTTAATGGACTACTATGGCATTAACACAGA	3180
Qy	3181	GGCTGAATACCTTAGTCGTGGCATTTATGAAGGATCAAAAACCTTTGACCGCAGAAAAA	3240
Db	3181	GGCTGAATACCTTAGTCGTGGCATTTATGAAGGATCAAAAACCTTTGACCGCAGAAAAA	3240
Qy	3241	TGCTGAGGCCATTAGTGTGCTGTGAGGCCCTTGAGGAAGGAGGCAAGAGCTGTTCAA	3300
Db	3241	TGCTGAGGCCATTAGTGTGCTGTGAGGCCCTTGAGGAAGGAGGCAAGAGCTGTTCAA	3300
Qy	3301	GAGGCGTAATCATATAGATGACATGTTTACCAAGGCTTCGGCTTGCTACCAAGTTACATA	3360
Db	3301	GAGGCGTAATCATATAGATGACATGTTTACCAAGGCTTCGGCTTGCTACCAAGTTACATA	3360



Db 301 GGTAAACAGTCCAAAGGAGGATCTAGAGCATTTGCCAAAGTTCAATTTGCCGACACAT 360  
QY 361 AAGTCTGACAAAATCATCACTTTGGCTAATAAACAGGCTGTATTTTGGCTCTCTCTTATTT 420  
Db 361 AAGTCTGACAAAATCATCACTTTGGCTAATAAACAGGCTGTATTTTGGCTCTCTCTTATTT 420  
QY 421 GAAGCTTTGGAAATGAACAGTATATTTGCCAAGTTCCTCAAGTTCGGGCAATATGTCATGATGGA 480  
Db 421 GAAGCTTTGGAAATGAACAGTATATTTGCCAAGTTCCTCAAGTTCGGGCAATATGTCATGATGGA 480  
QY 481 TGGCATAACTTTGAATTTCCGATGTCAGATATCAGATGACAAAGTTTGCAGTTGGGAAG 540  
Db 481 TGGCATAACTTTGAATTTCCGATGTCAGATATCAGATGACAAAGTTTGCAGTTGGGAAG 540  
QY 541 TACAGAAATTTCAATTCAAATTTGGCAATTTGGATGGAAGAAATTTTCTTTTATATCTAG 600  
Db 541 TACAGAAATTTCAATTCAAATTTGGCAATTTGGATGGAAGAAATTTTCTTTTATCTAG 600  
QY 601 TGGTTCAGTCACTATAAATTCAGCTTTCATATGCAAAATATATGGCAGGTTGTGCTCCA 660  
Db 601 TGGTTCAGTCACTATAAATTCAGCTTTCATATGCAAAATATATGGCAGGTTGTGCTCCA 660  
QY 661 TGGTCCATATGGTCAAAATGTCAGTTTCTCTCATACAGTTATTTGGTGTCTCTCGGAT 720  
Db 661 TGGTCCATATGGTCAAAATGTCAGTTTCTCTCATACAGTTATTTGGTGTCTCTCGGAT 720  
QY 721 CTATAGAGACTTGAAGAACTCTGTATAGCTTCTTTAAGGAAACTCTGTATGATCAGTG 780  
Db 721 CTATAGAGACTTGAAGAACTCTGTATAGCTTCTTTAAGGAAACTCTGTATGATCAGTG 780  
QY 781 GGTAGGACACACAGATTTCCCTCCATCTTGGATAGGGCTATCTCTAGCTTATTTTGA 840  
Db 781 GGTAGGACACACAGATTTCCCTCCATCTTGGATAGGGCTATCTCTAGCTTATTTTGA 840  
QY 841 GTTCGGTAGGGTGTCTCTTCCAAATTTTCAGAGAAAGTTTTCACATATGAGCAAGC 900  
Db 841 GTTCGGTAGGGTGTCTCTTCCAAATTTTCAGAGAAAGTTTTCACATATGAGCAAGC 900  
QY 901 TGAAGCAATATTAATTTACAGACTGTTTCCACCTTTTTCGTCCTCAAAATCGGCTCT 960  
Db 901 TGAAGCAATATTAATTTACAGACTGTTTCCACCTTTTTCGTCCTCAAAATCGGCTCT 960  
QY 961 GGTTCCTCAATGTCAGCTCCGAGGAAATTTCAATTTCCCTTACAGATTTTGTTCAAAT 1020  
Db 961 GGTTCCTCAATGTCAGCTCCGAGGAAATTTCAATTTCCCTTACAGATTTTGTTCAAAT 1020  
QY 1021 TAGTTCTTTTGTACAGCATGATGCATACCTGGCCAGCATTAATGTCATCTTTTCCG 1080  
Db 1021 TAGTTCTTTTGTACAGCATGATGCATACCTGGCCAGCATTAATGTCATCTTTTCCG 1080  
QY 1081 ATTAGTTGATCCTCGAAGGAGAAATGTGGCATGCAATGAGCATGCCCTAGAGAACTGTA 1140  
Db 1081 ATTAGTTGATCCTCGAAGGAGAAATGTGGCATGCAATGAGCATGCCCTAGAGAACTGTA 1140  
QY 1141 CTATATAAGGAGTCTGTTATGATCCGCTGAGGTGGCTCACTGAGCAGTATGATGGTA 1200  
Db 1141 CTATATAAGGAGTCTGTTATGATCCGCTGAGGTGGCTCACTGAGCAGTATGATGGTA 1200  
QY 1201 TCTCAAGGAGTACAACTCCAAATCTCCGTCATCACTTTAGATGATGGTGGTGTGTA 1260  
Db 1201 TCTCAAGGAGTACAACTCCAAATCTCCGTCATCACTTTAGATGATGGTGGTGTGTA 1260  
QY 1261 TGTAAAGAGGTCCTAGTAAACACCATGCAAGTTTATTTTGTGGTCCAGAGGTTAATGT 1320  
Db 1261 TGTAAAGAGGTCCTAGTAAACACCATGCAAGTTTATTTTGTGGTCCAGAGGTTAATGT 1320  
QY 1321 TTCCAAATCGGTTCTCCCAATTTATCTGGAAGACATAGATAACTTTCTCTGTTTCTCTT 1380  
Db 1321 TTCCAAATCGGTTCTCCCAATTTATCTGGAAGACATAGATAACTTTCTCTGTTTCTCTT 1380  
QY 1381 TGTGATGAGGAGTGGGAGAACTGTATCTACAGCTTATTACCAAGCAAGTACTGG 1440  
Db 1381 TGTGATGAGGAGTGGGAGAACTGTATCTACAGCTTATTACCAAGCAAGTACTGG 1440

QY 1441 AAGTGTGTGAGGACAAACATCTATGAGAGATCTTTATCAACTCTCGGAAAGCCTTTGT 1500  
Db 1441 AAGTGTGTGAGGACAAACATCTATGAGAGATCTTTATCAACTCTCGGAAAGCCTTTGT 1500  
QY 1501 AATTTGTGATATAAAATTTTGAATTTCTTGCATTTTCATCGAGCCAGTTGCGGATATTC 1560  
Db 1501 AATTTGTGATATAAAATTTTGAATTTCTTGCATTTTCATCGAGCCAGTTGCGGATATTC 1560  
QY 1561 AGTGTGATGTTTGCATCAAGACCTGGCCTTACTGCAATATGATATAGAGCTTGGATGG 1620  
Db 1561 AGTGTGATGTTTGCATCAAGACCTGGCCTTACTGCAATATGATATAGAGCTTGGATGG 1620  
QY 1621 TGATTTTCCAGATCAAGAAATGTCGAAATATGTCGAGACTTGGTCAATCTTTGG 1680  
Db 1621 TGATTTTCCAGATCAAGAAATGTCGAAATATGTCGAGACTTGGTCAATCTTTGG 1680  
QY 1681 TTCTCCAGAGAGACTTTGAGTGTCTTAGGCATGAGATTGAAGTTATTTCCCGATGTA 1740  
Db 1681 TTCTCCAGAGAGACTTTGAGTGTCTTAGGCATGAGATTGAAGTTATTTCCCGATGTA 1740  
QY 1741 GGTTCATGGAACAGCTATCTTTTCTGATGGAATTTGTAATATCTGGTGACCTTGC 1800  
Db 1741 GGTTCATGGAACAGCTATCTTTTCTGATGGAATTTGTAATATCTGGTGACCTTGC 1800  
QY 1801 TCATAGAGTTGCTCAAAATGTCGCTTCAATATACCCATCTGCTTCCAGATTCGTTA 1860  
Db 1801 TCATAGAGTTGCTCAAAATGTCGCTTCAATATACCCATCTGCTTCCAGATTCGTTA 1860  
QY 1861 TGTGTGATATAAGGTTGTTGCTGCTGATCCGATTCATCAATGAAGTTGCTTTGAG 1920  
Db 1861 TGTGTGATATAAGGTTGTTGCTGCTGATCCGATTCATCAATGAAGTTGCTTTGAG 1920  
QY 1921 AAGAGCATCTCGAATATGAATCAGACACATATAAGTTAGATGCTCTTGGATGGAGCAA 1980  
Db 1921 AAGAGCATCTCGAATATGAATCAGACACATATAAGTTAGATGCTCTTGGATGGAGCAA 1980  
QY 1981 ATATCAGCTTGTATCTTAACTGCTCAACTGATTACTCTCTGCTACACTTGGAGTGA 2040  
Db 1981 ATATCAGCTTGTATCTTAACTGCTCAACTGATTACTCTCTGCTACACTTGGAGTGA 2040  
QY 2041 AGATGAAGTTCTCGACAGAGCAAGAAAGAGCTTAGATCAGCTTGTATCTTGA 2100  
Db 2041 AGATGAAGTTCTCGACAGAGCAAGAAAGAGCTTAGATCAGCTTGTATCTTGA 2100  
QY 2101 TGATTTCTTCAAGGACACAGGAGCTTTGGAATTTGATGCTCTCTGAGAGAACTAATAT 2160  
Db 2101 TGATTTCTTCAAGGACACAGGAGCTTTGGAATTTGATGCTCTCTGAGAGAACTAATAT 2160  
QY 2161 TCTCAAGGCAATGCTAACTGCTGTTTAAAGCTGATGCTGAGCCCTTCTTTCAATGAT 2220  
Db 2161 TCTCAAGGCAATGCTAACTGCTGTTTAAAGCTGATGCTGAGCCCTTCTTTCAATGAT 2220  
QY 2221 GTTGCAAACTTCCGCGCATCCAAAGTTGCTGATTTGCGGACTAGATCAAGAAATTTTAT 2280  
Db 2221 GTTGCAAACTTCCGCGCATCCAAAGTTGCTGATTTGCGGACTAGATCAAGAAATTTTAT 2280  
QY 2281 TCCAAATGGAAGAACATGATGGATGTTTGGATGAATCCAGAACCTTGGAAATATGGTCA 2340  
Db 2281 TCCAAATGGAAGAACATGATGGATGTTTGGATGAATCCAGAACCTTGGAAATATGGTCA 2340  
QY 2341 GGTGTTTGTTCAGTTTACTGCTGACATGGAGAGTTTCTGAGAGTTTACATCCATT 2400  
Db 2341 GGTGTTTGTTCAGTTTACTGCTGACATGGAGAGTTTCTGAGAGTTTACATCCATT 2400  
QY 2401 TAATAACAGAGATCCCAACAGTAAATTTCTGAGGAAATGTGGTGTGTTGCAAA 2460  
Db 2401 TAATAACAGAGATCCCAACAGTAAATTTCTGAGGAAATGTGGTGTGTTGCAAA 2460  
QY 2461 AAATCCAGTGTGATCTGCTGATATTCGTTTAAAGGCTGTAATCTTCGAGGCT 2520  
Db 2461 AAATCCAGTGTGATCTGCTGATATTCGTTTAAAGGCTGTAATCTTCGAGGCT 2520





Db	318	AAAGGTGGGTTCCCAAGACGATTTGCAATATTCAATTCACACCGCAAAATCTGCTACA	377
Qy	371	AAAATCATCACTTTGGCTAATAAC-----AGCGTGTATTTTGGCTCTCTTATTATG	421
Db	378	TCATGATGTCAGAGCTAACAACATTTTGAGAACATTTGCGGTATGGGACCTCTATTTA	437
Qy	422	AAAGCTTGGGAATGAAAACACTGATTTGTCCAACTGC--GGGCATATGGGATCAGATG	478
Db	438	AAAGCTCGGGAAATGAAAAGAGATATGTGCCAAGCCCAAGGGTGTTTTGCATAGTTTG	497
Qy	479	GATGGCATCACTTTTGAATTTTCGGATGTCAGATATCAGATGACAAGTTTGCAGTGT--G	535
Db	498	GATGATGGAACGTCTTTTGGCTGTCAGATCTCAAGGGAAGNATCTCTGTGTTTATGG	557
Qy	536	GGAACTACAGAAGTTTCAATCAATTTGGCATTTGAGATGGAAGAAATTTTTTTCTTTT	595
Db	558	AAAAGCAGGATGTTATTGTAAATTTTGGAGTGGAAATGAGAAAGATGCAATTCCTTAT	617
Qy	596	TCATAGTGGTTACGCTCACTATAAATTCACGCTTTCATATGAAATATATGACAGTTGTG	655
Db	618	TCCCACAACTGTGCAATACAAACTGTGACCTTTCATATGAGACATTTTGGAAATTTG	677
Qy	656	CTCCATGCTCCATATAGTCAAAATGCTCAGTTTCTCCTCATACAGTTATTTTGGTCTCT	715
Db	678	CTGCATCGGCCACGGAATGAGACTACAGCTTATCTGTGATTTCAGTTACTTGGTCTCC	737
Qy	716	CGNATCTATAAGAC-----TTGAAACTCTCTGTATTAGC	751
Db	738	CGGTTTTTGGAGAACGATGTACCTACATCAACAANAATCTTTGTAGTACCTTTGTTCA	797
Qy	752	TTCTTTAAGGAAACTCCTGATGATCAGTGGGTGAGGACAAACAGATTTCCCTCC---	808
Db	798	TTCTTCAAGATGCCCTGATGAGCAATGATCCGAGCAATTGATTTCACTCCAGAAAGT	857
Qy	809	TGGATAGGCTATCTTCTAGCTTATGTTTGCAGTTTCCGTAGGGGTGTCGTCTTCCAAAT	868
Db	858	CGTATTGGGACGCTCCGCCATATGTCGTGAGCTTCCTAATGCGCCGCAACTTCCAAAT	917
Qy	869	TTGAGGAAAGTTTTTCCACATATGAGAACTGTGAAACAATATTACHTTACAGACTGGT	928
Db	918	TTCAGGAAAACTTTGCTTATATGAGGAAAGTGAGAGCAATACACTTACACACAGGA	977
Qy	929	TTACCTTTTTCGCTCTCAAAAACGGCTCTGGTTTCCCAATGCCAGCCTCCGAGAGGA	988
Db	978	GTTCCTCTT-----TCTCAAAATTTGGGCTCTGTCCCAATGTTGCTCTCTAGGT	1031
Qy	989	ATTTCAAATCCCTACAGATTTTGTTCAAAATTAGTCTTTGGTACACATGGATGCATA	1048
Db	1032	GTAAAAATCATATGACATCTTGTTTAAAGTCAATTCAATGGTTCAACATGCAATGCTT	1091
Qy	1049	CCTGGGCCAGCATTAATGCTCACTTTTCCGATTAGTTGATCTCTCGAAGGAGAAATGTG	1108
Db	1092	GCAGGACCTGCATTTGATGGTGACTCTATCCGCTGGTTGATCCACGTPAGAATGCCCGT	1151
Qy	1109	GCATGCATTGAGCATGCCCTAGNAACTGTACTATATAAGGAGTGCTGTATTGATCCC	1168
Db	1152	GAATTTATTGAATATGCTTTAGAAAAGATTTACTATTCAAGGAAATTTTGTATTGAACCC	1211
Qy	1169	GTGAGTGGCTCACTGAGCAGTATGATGGGTATCTCAAGGTTAGACAACCTCCAAAATCT	1228
Db	1212	ACAAAGTGGCTGACTGATCAGTACAAAACATACCTTGAGTCAAAAAATCATCCTCGGTCA	1271
Qy	1229	CCGTCCATCACTTTAGATGATGGTTGGTGTATGTAAAGAGGGTCTAGTAAACCAATGC	1288
Db	1272	CCTGCAATATCCTTGGATACAGGGTTGGTATACGTTCCGACGGTTTCAGATCACGCCCTGC	1331
Qy	1289	AAAGTTTATTTTGTGTCACAGGTTAATGTTTCCAAATCGGGTTCTCCGCAAAATTTCT	1348
Db	1332	AAAGTATACTTTTGTGGTCCAGAGATGATGCTCAAACTGTTGTTCTCCGTCATTTCGGT	1391
Qy	1349	GAAGACATAGATAACTTTCTCGTGTTCCTTTTGTGTGATGAGAGTGGGAGAACTGTAT	1408
Db	1392	GAACATATTGATAACTTTCTAGCTGTTTCAATTTGTTGATGAAGAAATGGATAAACTGTTT	1451

```
QY 2477 CCTGTGATATTCGTGTTTAAAGGCTGTAATCTTCGAGCGCTGCACACATGGTAGAT 2536
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2511 CCAGGTGATGCGGTGTTTACAGCTGTGGAATGCCAGATTTGTACCACATGGTGGAC 2570

QY 2537 TGTGTTGATTCCTTCAGAAAGAAAAGACCTCATCCGAATGAATTTCTGGGAGTGAT 2596
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2571 TGTGTTGTTTCCCTCAAAAAGGACCAAGACCTCATCCAATGAGTTCTGGGAAGTGAT 2630
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2597 TTGATGGGATATCTACTTGTGTTGGGATCAAGACATGATCCCGCCAAAGGCAAGTC 2656
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2631 CTGATGGAGATCTACTTGTGTTGGGACCATGAATGATTCCTTCGCCCAATT 2690
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2657 CAGCCGATGGAATATCCTCCAGCACCCAGCATACAGTTGGACCATGATGCACAAATGAG 2716
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2691 GATCCAATGGACTATCTGCTCCGCACTGTGGAATTTGGATCATGATGTGATGATCGAG 2750
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2717 GAAGTTGAAGAGTACTTACCACCACTATATGTGTAATCAGACATTTGGGAATCATAGCAAT 2776
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2751 GAGTTGAGGAGTATTTGCCAATACATAGTCAATGACAGTCTGGGAATTAATGGCCAA 2810
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2777 GCCCATGTCGTATTTGCAGACAGAGAACCTCGATATAGCCATGAGTATCCATGCCAAAAA 2836
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2811 GCACACACTGTCTTGGCAGATAAAGAACATTTGAAGCAATGTCTGATCAATGTGTTAAG 2870
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2837 CTTCGTGAGCTCTTCAATTCGAGTGGACTTTCCAAAGACTGTTCCCGCTGGAATA 2896
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2871 CTTCGAAGTTGTTTCAACAGCAGTTGACTTTCCCTAAAACCTGGTGTCCAGCTGTTATA 2930
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2897 CCATCTCAGTTGCGCCCTAAAGATATCCAGACTTCATGATAGCGGACAAAGACCAGC 2956
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2931 CCTCTGAACCTTATGTCAAAGATATCTTGACTTCATGGAAGCCTGCACAAACCCACA 2990
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2957 TATATCTCAGAAAGAGTATTTGGAAGCTTTTCAGGAAGTGAAGGACAAAGACCTCAG 3016
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2991 TACAATTCGCAATACGTTGATAGGAAGCTTTTAGGAAGTGAAGAAATATCAACAAGT 3050
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3017 GCTAGCTCTATCGGACCTTCACAAGAGATGTTTCAAGAGGAGTCAATATGCTGATG 3076
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3051 GCCGCTCAATATACCTCTCACAAGTGTGTCGAGAGACTTTAGACCATGAATG 3110
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3077 GAAGTTGATGATTTGAAGTATCATTTGACGAGCTTTTGACTACAAAACCTGAATATGAC 3136
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3111 GAAGTTGATGCTTCATGGATTATGTTGATGATGCTTTCTATCACAACCAATATATGAC 3170
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3137 AACAGCTGGGTAAATTAATGGCTACTATGGCATAAACACAGAGGCTGAAATACTTAGT 3196
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3171 TACAAGTTGGGAATCTGATGGACTATATGGGATCAAAACTGAAGCTGAATCTCCGT 3230
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3197 GGTGCAATTAAGAGGCATCAAAAACCTTTTGACCGCAGAAAAGATGCTGAGGCCATTAGT 3256
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3231 GGGATATTTAGAAATGTCAAAATCTTCAACAAAAGGAGGGATGCAGAAGCAATCAAT 3290
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3257 GTTGTGTGAGGCTTCAGGAGGAGGCGAAGACCTGGTTC-----AAGAGCGTAAT 3310
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3291 ATGGCTGAGGTCCTTAAGAAAGAGGCGCAGGCGCTTGTTTCAATGAACACAGCGTGGT 3350
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3311 GATATAGA-----TGACATGTTACAAAAGCCTTCGGCTTGGTACCACGTTTACA 3358
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3351 GATGATGATTCAGGAGTAGTGATGTATGTCAAAAGCTTCTGCTTGGTACCATGTTACT 3410
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3359 TATCATCTACATATGGGTTGTACAAATCAGGGTTGAAAGAGCTCATTTCAATGAC 3418
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3411 TATCATCAAGTTACTGGGTTGTATTAATGAAGGCATGAATAGGGATCAATATCTAAGT 3470
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3419 TTCCCTCGTGTGTTATGACACACTAATCCAGATTAAGAGGACAAA 3466
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3471 TTCATGTTGTTTACCCCTCTTGTTCCTCAATCAAGAAAGAGAAA 3518
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 5

AAC63739

ID AAC63739 standard; cDNA; 3737 BP.

XX

```
AC AAC63739;
XX 07-FEB-2001 (first entry)
DT
XX
DE Maize RNA-directed RNA polymerase EST sequence, SEQ ID NO: 5.
XX
KW Maize; plant; RNA-directed RNA polymerase; expressed sequence tag; EST;
KW gene mapping; gene marker; plant virus resistance; plant breeding; ss.
XX
OS Zea mays.
XX
PN WO200060097-A1.
XX
PD 12-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US09105.
XX
PR 07-APR-1999; 99US-0128094.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;
XX
WPI: 2000-679376/66.
DR P-PSDB; AAB28330.
XX
PT New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT for controlling gene expression and providing mechanisms to engineer
PT plant virus resistance.
XX
PS Claim 2; Page 35-36; 62pp; English.
XX
CC The present sequence is one of a number of cDNA molecules which encode
CC plant RNA-directed RNA polymerase proteins. The sequences are useful as
CC probes for genetically and physically mapping genes, and as markers for
CC traits linked to those genes. They are useful for controlling gene
CC expression and provide mechanisms to engineer plant virus resistance.
CC They are also useful for plant breeding to develop lines with desired
CC phenotypes.
XX
SQ Sequence 3737 BP; 1063 A; 760 C; 863 G; 1051 T; 0 other;

Query Match 26.5%; Score 987.8; DB 21; Length 3737;
Best Local Similarity 59.8%; Pred. No. 7e-261;
Matches 1997; Conservative 0; Mismatches 1207; Indels 135; Gaps 15;

QY 195 TGGGAAGCAATTCAGGTTTCGGATTCCCTTATCTCTCTCTCGGGAAGTGGTTAAT 254
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 TTGGGAGAACTATTCAGGTCCTCAAGGTTTGTCTCACTGACAGTCCGGAATCTGCAAT 143
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 255 CATCTCTAGAGAAATATACAGGATATGGAACCTGATGTGCATTGGAGTTAAACAGTCCA 314
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144 TGTTTTGGAGCAATTCGTTGCTGGAACCATCTGTCTCTCAAGCTCAGGCATCCAA 203
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 315 AA-----GGAGATCTAGAGCAATTTGCCAAGTTCAATTTGCCGACACATAAGTG 365
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 204 GGAACATCTCTGCCAACTCAAGGCAATTTGCTATAGTTCAAGTCCAGTCCAGGAAAGTG 263
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 366 CTGACAAAATCATCTCTTGGCTAATA---ACAGGCTGTATTTTGGCTCTCTCTATTGCA 422
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 264 CTTCATTTGTAGAGATGCGCTCAAGACAGGTTCTCAGATTTGACGTTTATCTGA 323
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 423 AGGCTTTGGAAATGAAAACCTGATATTGTCC---ACTCGGGGCATATGTGGATCAGATGG 479
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 324 GAACACAGACCTGCAGACCGGACATTTGTCCAAGACCAAGGATTTCCAATGTTTCTCTAG 383
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 480 ATGGCATAACTTTGATTTCCGGATGTCAG--ATATCATGATCACAAGTTTGCAGTGTGG 536
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 384 AGGACATTTGCTGCAATTTGGGATGCTTGGTTAAGAAATATCCTATCTGCTCTTTTA 443
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 537 GAAGTACAGAAGTTTCAATTCATTTGGCATTGGATTGAAGAAATTTTTTCTTTTAT 596
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 444 GAGCAAGTAATGTTTCGGTTCAATTTGGATTGATATGAAAAAGATCTACTTCTACCTCT 503  
Qy 597 CTAGTGGTTCAGCTAGCTATPAAACTTCAGCTTCATATGAAATATATGCGAGTTGCG 656  
Db 504 CCTCAATTTTACATAATTTAACTTTGAACCTTCTTACGAAAGTATATGGAGATGCGC 563  
Qy 657 TCCATCGTCCATATGGTCAAAATGCT-----CAGTTTCTCCCTCATACAGTATTTGGTG 710  
Db 564 TTCACCGTCCACCTGCTTATAGGTCACGGACACAGTTCCTTTTGATTCAAGTTCAGGCAG 623  
Qy 711 CTCCTCGGATCTATAAGAGAC-----TTGAAACTCCTGTT 746  
Db 624 CTCCTAAATTTATAAACTGCTCCACGCGCTCCAGGTCFTATGTTTGGAGATCCTTTCT 683  
Qy 747 ATAGCTTCTTTAAGGAAACTCCTGATGATCAGTGGGTGAGGACAAACAGATTTCCCTCCAT 806  
Db 684 TCACTGGTTTAGGATGACACAGATGAACAATGACACAGACAATTTGATTTACTCCAT 743  
Qy 807 ---CTTGATAGGGTATCTCTAGCTTATGTTTGGAGTTCCTGAGGGTGTCTGCTTTC 863  
Db 744 CAGCTAGCATCGGCAATCATCTATTTATGCTCGAGGTGCCACACAGATGTGAGGCTT 803  
Qy 864 CAATTTCCAGGAAAGTTTTTCCACTATGACAGAACGTGAAACAAATATTACTTTACAGA 923  
Db 804 CAAGNAITGGCGACTACTTTTGTACTATTAAGAGCAGAACTTTGACTTTGAATGTCGA 863  
Qy 924 CTGPTTCACCTTTTCTGCTCTCAAAATCGGCTCTGGTTCCTCAATGTCGAGCTCCGG 983  
Db 864 ATGGGTATTATCTCTGTGTGTAGCAA-----CCTGTGTACCAATTTGAAATCTCCTG 917  
Qy 984 AAGGAATTTCAATCCCTACAGATTTTGTCAAAATTTAGTTCTTTGGTACAGCATGGAT 1043  
Db 918 ATTACATAGAGTCCCTTATGAGATACTTCTCAAAATCAACCAATTTGGTTCAGATGGGA 977  
Qy 1044 GCATACCTGGGCGCAGCAATAAATGCTACTTTTCCGATTTAGTTGATCTCGAAGGAGAA 1103  
Db 978 CACTCAGTGGGCGCAACAGTGTATAGTTTCTTCCGTCATGTTAGCCCAAAATTTGAAC 1037  
Qy 1104 ATGTGCGATGATTCAGATGCTTATAGAGAACTGTACTATATAAGGAGTGTCTGTATG 1163  
Db 1038 CTATTGATCATATAAAGCAGCACTTTTAAGATGACATATTTGAAAGCAGCTGCTGA 1097  
Qy 1164 ATCCCGTGAAGGTGCTACTGAGCAGTATGATGGGTATCTCAAGGGTATAGA---ACCTC 1220  
Db 1098 ACCCAACAGATGGTTATCTGTGCAATACTCCAGAAATACGGAATACGCCATGCAATCAC 1157  
Qy 1221 CAAATCTCCGTCACATCTAGATGATGGTGGTGTATGTAAGAGGGTCTCTAGTAA 1280  
Db 1158 AAAAGTTATCTAATATATCTCTGGATGATGGCTTGGTCTATGTCACAGGGTGCAGTTA 1217  
Qy 1281 CACCATGCAAAAGTTTATTTTGTGTCAGAGGTTAATGTTTCCAAATCGGGTCTCCGCA 1340  
Db 1218 CCCCTGCTAAAGTGTATTTTATGGACCTGAGATAAATGCTCCAAATCGCGTTGTCGCG 1277  
Qy 1341 ATTATCTGAAGACATAGATAACTTTCTTCTGTTTCTTTTGTGATGAGGAGTGGGAGA 1400  
Db 1278 ATTCTCTCGACACATAGATAACTTCTCTGGGATTTCAATTTGTTGATGAAGACTGTGGA 1337  
Qy 1401 AACTGTATTTACAGACTTATTACCAAAAGCAAGTACTGGAAAGTGGTCTCAGGACAAACA 1460  
Db 1338 AGCTCGTTACAGTGTATGCTACCTCGATCTACTCTCGAAATGATCAAGGAGAACTG 1397  
Qy 1461 ---TCTATGAGAGGATCTTATCAACTCTGCGGAAAGGCTTTGTAATGGTGTATAAAAAAT 1517  
Db 1398 CTCTGTATAATAGAGTTTGTGAGTCTTCAAAATGGCATCAATATTGGTGCAGAGCACT 1457  
Qy 1518 TTGAATTTCTGCAATTTCTATCGCCAGTTCGGGATAAATTCAGTGTGATGTTTGGAT 1577  
Db 1458 TTGAGTTTCTTGTCTTTCTCAAGTCAGCTTCGAGATAACTCTGCAATGATGTTGCTT 1517  
Qy 1578 CAAGACCTGGCTTACTCTCAATGATATAGAGCTTGGATGGGTGATTTTTCGCGAGATCA 1637  
Db 1518 CTCGGCAGGATGATCTGCGAGCGACATAAGGAAGTGGATGGGACCTTCGAGATATCA 1577

Qy 1638 AGAATGTCGCAAAATATGCTGCCAGACTTGGTCAATCTTTTGGTTCTCTCCAGAGACATT 1697  
Db 1578 GAAATGTCGCAAAATATGCTGCAATCTTTTGGGCAATCTTTTCAAGTCTCTTCTTCT 1637  
Qy 1698 TGAGTGTCTTAGGATGAGATGAAGTATTTCCCGATGAAAGGTTCAATGGAACAGCT 1757  
Db 1638 TAAAGTACACAAATCTGAGGTGGAAGAAATCTCTGATATTA---CAATGCGCAAAAGT 1694  
Qy 1758 ATGCTCTTTCTGATGAAATTTGTAATATCTGGTGACTTTGCTCATAGAGTTGCTCAA 1817  
Db 1695 ACATATCTCTGATGAGTGGAAAGATCTACGTAATTTTGCAGTGGAGTGGCTATGA 1754  
Qy 1818 AATGTCGCTTCAA---TATACCCCTCTGCTTTCAGATTCGTTATGTTGATATAAG 1874  
Db 1755 AGTSCAAATTTGAACGCTTTGCTCTTCTTCTCAGATGAGGTATGGCGTTACAGAG 1814  
Qy 1875 GTCTGTGGGTGTGATCGGATTCATCAATGAAGTTGCTTTTGAGAAAGAGATGTCGA 1934  
Db 1815 GTGTTGCTGTAGATACAGATCAATCATTAAGCTTTCTTTGAGAAAAGCATGTCGA 1874  
Qy 1935 AATATGAATCAGACACATAAAGTTAGATGCTCTTGGATGGAGCAATATATCAGCTTGT 1994  
Db 1875 AGTTCAGTCAGAAAATATCACTCTTGAIGTCTCTACACAGCAAGTACCAACCATGCT 1934  
Qy 1995 ATCTTAATGCTCAACTGATTTACGCTCTTGTCTACACTTGGAGTGAAGATGAAGTTCTCG 2054  
Db 1935 TCTGTAATCGGAGTTGATTTACTTCTCTCAGACTTGGGTTAGCGATATGTTCTTGG 1994  
Qy 2055 AACAGAAGCAAAAGGAGCTGTAGATCAGCTGTATGCTATCTTTCATGATCTTTTGAAG 2114  
Db 1995 AGCTAAAGCAGAGGAGGCTTAAAGCAGTTGAACAGATGTAACGTGAACACACAGGCTG 2054  
Qy 2115 CACAGGAGGCTTGGAAATGATGCTCTCTCGAGAGAACACATAATATTCTCAAGGCAATGC 2174  
Db 2055 CTGTAAGCAGTTGAATTTATGCCCCATGGAGAGGTAAACATGTAGTTTAAAGAAATGT 2114  
Qy 2175 TAAACTGTGTTTAAAGCTGTAGCTGAGCCCTTTCTTCTCAATGATGTTTCAACCTTCC 2234  
Db 2115 TGTCATGTGCTACAGCTGATCATGAGCCATCTTTTCCATGCTGCTCAAACTTTTA 2174  
Qy 2235 GCGCATCAAGTTGCTGATTTGGGACTAGATCAAGAATATTATTCCAAATGGAGAA 2294  
Db 2175 GAGATCCAAAGCTCTAGATTTGAAACAAGTCAAGGATATTCATCACACAGGGGAG 2234  
Qy 2295 CAATGATGGATGTTTGGATGAATCCAGAACCTTGGAAATGTTGTCAGGTTTGTTCAGT 2354  
Db 2235 CAATGATGGTTGCTGGATGAAACCTGCACTTAAAGTACGGCCAGGTATTCGTCOAAG 2294  
Qy 2355 TTACTGTGCTGGACATGGAGAGTTTCTGACGATTTACATCCATTTAATAACAGCAGAT 2414  
Db 2295 CTCTTTACAGTGCAGATG----- 2312  
Qy 2415 CCACCAACAGTAAATTTCAATCTGAAGGGAATGTTGTTGTCAAAAAATCCATGCTTGC 2474  
Db 2313 ---ACCATCCGCAAGTCTGTTAACTGGAAGAGTGTGTCGCCAAAAAATCTTGTCTCC 2369  
Qy 2475 ATCTGTTGATATTCGTTTAAAGCTGTAAATGTTGAGGCTGCACACATGTFAG 2534  
Db 2370 ACCGTGTGATACAGGTTCTCCAGGCTGTGATGTTCTGCTCTGCACCCCTGTTG 2429  
Qy 2535 ATTGTGTTGATTCCTCTCAGAAAGGAAAAAGACCTCATCCGAATGAATGTTCTGGGAGT 2594  
Db 2430 ACTGTGTTGCTTTCCACAGCAGGACCAAGCCGACCCCTAATGAGTGTTCAGGAGT 2489  
Qy 2595 ATTTGGATGGGATATATCTATTTGTTGCTGGGATCAAGACATGATCCCGCCAGGCAAG 2654  
Db 2490 ATCTGATGGGACATATATTTTCTTGGGATCCACATCTTATTCCAAAGTCTTGG 2549  
Qy 2655 TCCAGCCGATGAATCTCTCCAGCCACGACATACAGATGTCGACCATGATGTCACATTTG 2714  
Db 2550 TGGATCCTATGACTATACTCCAGCTTCAGCAGAAACATFAGACCATGATGTCATTTG 2609

```
QY 2715 AGGAAGTTGAGAGTACTTCCACCAACTATATTGTGATGACAGTGTGGGAATCATAGCAA 2774
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 2610 AGGAGATACAGGAGTACTTCCAAACATACATAGTAAATGAGAGTCTTGGGATATCGCCA 2669
QY 2775 ATGCCCATGTCTGATTTTGCACACAGAGAAGCTGATATGGCCATGATGATCCCATGCAAAA 2834
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 2670 ATGGCATGTGCTTTTACAGATCAGCAACGATGATGAAAGCTGAGAGTCCACCGTGGCTC 2729
QY 2835 AACTTGTGAGCTCTTTTCAATTCAGTGCAGTTCCTCAAGACTGGTGTCCCGCTGAAA 2894
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 2730 AACTGGCCAAAGCTCTTCTATAGCTGTGATTTCCCAAGACTGGAGTCCCGCTCTGA 2789
QY 2895 TACCATCTCAGTTCGCCCTTAAAGAAATACCAGACTTTCATGTAAGCCGAGCAAGACCA 2954
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 2790 TTCCACATGAGCTACATGTCAGAGATCTCTGACTTCATGGAGAACTCGACAAAGTCA 2849
QY 2955 GCTATATCTCAGAAGAGTATTTCGAAGCTTTTCAGGAAGTGAAGCAAAAGCACCTC 3014
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 2850 CCTATGAATCAAGAGGTGTATCGGGAAGCTCTATAGGAAATAAAGAACACACACAC 2909
QY 3015 AGGCTAGCTCTATCGGACCTTCCACAGAGATGTTGCAAGGAGATCATATGATGCTGATA 3074
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 2910 A-----CATAAAGCACTTCACGAGGAGTGGCAAGCGCTCTATGACACCGATT 2960
QY 3075 TGAAGTTGATGGATTGAGATACATGACGAAGCTTTTGACTACAAACTGATATG 3134
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 2961 TGATTTGATGGCTATGAGATTACATTTACTGAGGCTATAGATTCAGGAAGAGTACG 3020
QY 3135 ACAACAGCTGGTAATTTAATGACTACTATGTCATAAAACAGAGGCTGAAATACTTA 3194
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 3021 ATTTGAGCTGGTAATCTTATGACCACTATGTCATGAAAGTGAAGTGAAGTGAAGTAA 3080
QY 3195 GTGTTGCAATATGAAGCATCAAAAACCTTTTGACCGCAAGAAAGATGCTGAGGCCATTA 3254
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 3081 GTGGATGATTTCTAAGATGCAAGAAATTTCCACAAAGATGATGATGCTGATCAATTA 3140
QY 3255 GTGTTGCTGTAGGCCCTTGGAGGAGGAGCAAGAGCTGTTTCAAGAGCGGTA----- 3308
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 3141 GAATGCGGTGAGATCTTTGAGGAAGAAAGCTAGGTCTGTTGTTCAATGAGTAGGACAG 3200
QY 3309 -----ATGATATAGATGACATGTTTACCAAGAGGCTTCGGCTTGTACACCTTACATATC 3362
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 3201 GAGAGGATGGCAAGATGCCATGGAGGCCAAGGCTCTGCTTGTACCATGTTACTTATC 3260
QY 3363 ATCCATACATATGGGGTGTCTACATCAAGGCTGGAAGAGCTCATTTCAATGAGCTTTC 3422
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 3261 ATCAGCAGTACTGGGCGAGCTACATGAAGGCTATGATCGGCCGATCTTTATGCTTCC 3320
QY 3423 CTGTGTGTTTATGACAGCTAATCCAGATTAAGAGG 3461
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 3321 CATGTGGCTATGACAAGCTTGTGGCCATCAGCAGG 3359

RESULT 6
AAC63740
ID AAC63740 standard; cDNA: 3901 BP.
XX
AC AAC63740;
XX
XX 07-FEB-2001 (first entry)
XX
DE Maize RNA-directed RNA polymerase FIS sequence, SEQ ID NO: 7.
XX
KW Maize; plant: RNA-directed RNA polymerase; FIS: gene mapping;
KW gene marker; plant virus resistance; plant breeding; ss.
XX
OS Zea mays.
XX
XX WO2000060097-A1.
XX
XX 12-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US09105.
XX
PF
```

```
XX 07-APR-1999; 99US-0128094.
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX (PION-) PIONEER HI-BRED INT INC.
PA
XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;
XX WPI: 2000-679376/66.
XX P-PSDB; AAB28531.
XX
XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
XX PT for controlling gene expression and providing mechanisms to engineer
XX PT plant virus resistance -
XX
XX Claim 2; Page 40-41; 62pp; English.
XX
XX The present sequence is one of a number of cDNA molecules which encode
XX plant RNA-directed RNA polymerase proteins. The sequences are useful as
XX probes for genetically and physically mapping genes, and as markers for
XX traits linked to those genes. They are useful for controlling gene
XX expression and provide mechanisms to engineer plant virus resistance.
XX They are also useful for plant breeding to develop lines with desired
XX phenotypes.
XX
XX Sequence 3901 BP; 928 A; 975 C; 1058 G; 935 T; 5 other;
XX
XX Query Match 9.2%; Score 345; DB 21; Length 3901;
XX Best Local Similarity 52.4%; Pred. No. 5.4e-84;
XX Matches 1049; Conservative 0; Mismatches 900; Indels 54; Gaps 11;
QY 1223 AATCTCCGTCCTACCTTTAGATGATGGTGTGTTATGTAAGAGGCTCCTAGTAACA 1282
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1290 AAACATCTTCGCAACAGAGTGGGATGTCATCTCTGAGTGAGGAGGCTGTAATCAGC 1349
QY 1283 CCATGCAAGATTTATTTTGTGTCAGAGAGTTAATGTTTCCAATCGGTTTCGCCAAT 1342
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1350 CCCACCAAGCTTTATGTCATGCCACCAAGAGTGGAGCGCTCTAATCGCTCATCCGCCAT 1409
QY 1343 TATCTGACACATACATACATTTCTGTTGTTCTTTGTTGATCAGGAGTGGAGA-- 1400
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1410 TATAGTGAAGTCTCAGACCGGTTTCTGAGGTTTACTTTATGGATGGGAATGACAGATG 1469
QY 1401 ---AAGTGTATCTACAGACTTATTACCAAGCAAGTACTGGAAGTGTGTG----- 1450
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1470 CTCACAGATGATGCTGTAATTTCTGCTGCACAAATGCTCAAGATTTGATGTCAAAC 1529
QY 1451 -----AGGCAACATCTATGAGAGATCTTATCACTCTCGGAAAGCTTT 1498
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1530 TCGTTCTGCTAAGACACAGGTACAGCGGTGTTAAACGTTTTCAGAGGAGTTC 1589
QY 1499 GTATTGGTGATAAAAAATTTGAATTTCTTTCATTCGAGCCAGTTGCGGAGATAAT 1558
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1590 CACATGTGTGGCAGGAGTACTCGTTCTTTCATCTCACTTAACCAAGCTGAGGACAGG 1649
QY 1559 TCAGTGTGATGTTTGCATCAAGACCTGGCTTACTGCAATGATATAGAGCTTGGATG 1618
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1650 TCAGCATGTTCTTCGACAGGACAGAACACACAGTGGAAACCACTTAGAAATGATG 1709
QY 1619 GGTGATTTTCGCAGATCAAGAAATGTCGCAAAATATGCTGCCAGACTTGTCTAATCTTT 1678
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1710 GGGCGGTTCAC---AGTAAGATGTAAGCAAGCATCGCGCTCGATGGGCGAGTCTTC 1766
QY 1679 GGTTCCTCCAGAGAGACTTTGAGTGTCTTTAGGCATGAGATTGAAGTATTCGCCGATGA 1738
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1767 TCCTCTACATATGCTACGGTGGTCTGACGCCGATGAGGTAATGAGTCTCTTGATGAA 1826
QY 1739 AAGTTTCATGAACACGACTATGCTTTTCTGATGGAATGTTGTAATAATATCTGTGACTTT 1798
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
PA 1827 GTTGAACAT---AACGGGTACATTTTCTCTGATGGAATGTCGCAAGATTACGTGGACCTT 1883
QY 1799 GTCATAGAGTTGCCTCAAAATGTGGCTTCAATATA---CCCCATGCTGTTTCCAGATT 1855
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 1884 GCACCTGGAAGTGTCTCAAGAGCTCCAAATTCACAGATAATCCCCATCTGCTTACCAGATT 1943  
Qy 1856 CGTTATGGTGGATATAAAGGTGTGTGGGTGTGATCCGG-----ATTCAATCAATGAAG 1909  
Db 1944 AGGTATGAGGCTTCAAGGGTGTATATCTCTGCTGGGAAGAAATAATGATGGATACGA 2003  
Qy 1910 TTGTCTTTGAGAAAGACATGTCGAATATGAATACACACACATAAAGTTAGATGTCCTT 1969  
Db 2004 CTTTCCCTGAGCGCGAGCATGCACAAGTTTGAAGTCTAACCATATCTGTTAGAGTGGTC 2063  
Qy 1970 GGATGGAGCAATATCAGCCCTTGTATCTTAATCGTCAACATGATAGCGCTCTTGTCPACA 2029  
Db 2064 TCGTGGACAAAGTTTACCGCAGGATCTTAATCGTCAGATTAATTACATTACTCTCTCCTC 2123  
Qy 2030 CTTGGAGTGAAGATGAAGTCTCGACAGAGCAAGCAAAAGAGAGCTGAGATCAGCTTGAT 2089  
Db 2124 TTGAATGCCCGGATGCTATCTTTCCTCAAAATGCGAGAGCCATGTTATCTTAATCTCAAC 2183  
Qy 2090 GCTATCTTGCATGATCTTTTGAAGGCACAGGAGGCTTTGGAATTCATGCTCTCTCGAGAG 2149  
Db 2184 AATATTTTGCAGACTCTGATGTTGCTTTTGA---CATTTGAACCGCTCTTGTGCTGAG 2240  
Qy 2150 AACACTAATATCTCAAGGCAATGCTAAACHTGCTGTTATAGCTGATGCTGAGCCCTTTT 2209  
Db 2241 CAAGGAACCACTGCGACACTGATGTGTAGTCTGCTGCAATTCACCTGGAACCTGAGCCACAC 2300  
Qy 2210 CTTTCAATGATGTTGCCAAACCTTCGCGCATCCAAAGTGTCTGATTTGCGGACTAGATCA 2269  
Db 2301 CTGAAGCAATGCTGTTAGCTATAGGTCCTCACAGCTGCTAGTCTTTTGGAGAAGACA 2360  
Qy 2270 AGAATATTTTCCAAATGGAAGAACAAATGATGGATGTTTGGATGAATCCAGAACCTTG 2329  
Db 2361 AGGATTTTGTCCCAAGGAGAGTGGTGTGATGGCTGCCCTGATGAATCTGGGATCCCT 2420  
Qy 2330 GAATATGTCAGGTGTTGTTTTCAGTTTACTGCTGTGACATGAGAGTGTCTGACGAT 2389  
Db 2421 GAGCAGGCACTGCTTATTCGCGGCATCACTCCATCAC-TCAATATTCGTGCTGAA 2479  
Qy 2390 TTACATCCATTTAATACAGAGATCCACCAACATAATTTCAATCTCGAAGGAAATGTG 2449  
Db 2480 GTATGATCAAGATTGCTGTCAG--CAAAACACAAATGAGAGACCATCTCGGTACTATC 2537  
Qy 2450 GTTGTGTCAAAAATTCATGCTGTGATCTGCTGATATTCGTGTTTAAAGGCTGTAAT 2509  
Db 2538 GTAAATGGCAAGATCCATGCCCTTCATCCAGGGATGCGGAATCCTTGAAGCTTTGAT 2597  
Qy 2510 GTTCGAGCGCTGCACACATGATGTTGTTATTCCTTCAGAAAGGAAAAAGACCT 2569  
Db 2598 GTCCCTGAACATGATCACCTTGTGATGCTTGTGCTTCCCAAGAAAGGTGAGAGGCG 2657  
Qy 2570 CATCCGAATGAATGTTCTGGGAGTGAATTTGGATGGGATATCTACTTTGTTGCTGGGAT 2629  
Db 2658 CAGCGGAATGAAGCATCTGGGAGTATCTTGTATGGGATCTATCTTGTAAACATGGGAT 2717  
Qy 2630 CAAGACATGATCCGCGC-----AAGGCAAGTCCAGCCGATGAATATCCTCCAGCACCC 2683  
Db 2718 GAAACCTTATACCACCTGGTAAAGAGATTGGAAACCTTATGACTACTCCCGACGTGAA 2777  
Qy 2684 AGCATACAGTTGGAACATGATGTCAAATTTAGGAAGTTGAAGTACTTTCACCAACTAT 2743  
Db 2778 GCAAAACAACTGCCACGCGAGTATCCCAACATGATATTTGTTGTTCTTCTTGAAGAAC 2837  
Qy 2744 ATTGTGAATGACAGTTTGGGAATCATAGCAATGCGCATGCTGTAATTTTCAGACAGAGAA 2803  
Db 2838 ATGTAATAGGAACATGGTCCAAATAGCAATGCTCATGTTGTTCACTGATATGAGC 2897  
Qy 2804 CCGATATGCCATGATGATCATCATCAAAAACTTGTGAGCTCTTTCAATTTGAGTG 2863  
Db 2898 GAGTATGAGCAATGATGAGAGTGTATTCAGTTGGCAGAACTAGCAGCAACTGCTGTG 2957  
Qy 2864 GACTTTCCAAAGACTGGTGTTCGCTGAAATACCATCTCAGTTGCGCCCTTAAGATAC 2923  
Db 2958 GACTTCCCAAGACAGCAAAATTTGTCAATGCCAGCATTCCTTCGACCAAAATTTATAT 3017

Qy 2924 CCAGACTTCATGATAAGCCGAGACACAGCTATATCTCAGAAAGAGTATTATGGAAG 2983  
Db 3018 CTTGACTTCATGGGAAGGAGGATGCTATCTCTTAGATCAGAGAAGATCTTGAAGG 3077  
Qy 2984 CTTTTCAGAAAGTGAAGGACAAACACCTCAGGCTAGCTCTATCCGACCTTCAACA 3043  
Db 3078 CTTTATCGGTCAATCCAAG---AAGCCTCCAGCGATGATTGGTTCCAGAAAGAACTTGC 3134  
Qy 3044 GATGTTGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGATTGAAGATTACATT 3103  
Db 3135 ACATCTAAACAATCTGCTTATGATGAGATATGGAAGTTGCTGGTGCAGCTGATTTC 3194  
Qy 3104 GAGCAAGCTTTTGACTACAAACTGAATATGACAAACAGCTGGTATTAATTAATGGA 3163  
Db 3195 TCGAGTGTCTGGCAGTGCAGTCTCATATGAAACACAACTGAACGCACTGCTCAACCA 3254  
Qy 3164 TATGGCATATAAAACAGAGCTGA 3186  
Db 3255 TATGGCGTGGCAGCTGAAGCAGA 3277

## RESULT 7

AAH77704  
ID AAH77704 standard; cDNA; 3591 BP.  
XX  
AC AAH77704;  
XX AC 1. 3591  
DT DT /\*tag= a  
XX 13-NOV-2001 (first entry) /product= "SGS2"  
XX Nucleotide sequence of a plant SGS2 polypeptide.  
XX  
XX SGS2; RNA-dependent RNA polymerase; transgene silencing;  
KW transgene stability; crop plant; viral resistance; ss.  
XX Arabidopsis thaliana.

Key Location/Qualifiers  
CDS 1. 3591  
FT /\*tag= a  
FT /product= "SGS2"

FR2804128-A1.

27-JUL-2001.

26-JAN-2000; 2000FR-0001007.

26-JAN-2000; 2000FR-0001007.

(RHOB-) RHOBIO.  
(INRG) INST NAT RECH AGRONOMIQUE.

Beclin C, Elmavan T, Mourrain P, Vaucheret H;

WPI; 2001-543303/61.

P-PSDB; AAG67235.

A new plant SGS2 gene involved in encoding an RNA-dependent RNA polymerase and in transgene silencing, increases transgene stability and expression in transgenic plants when it is inactivated -

Claim 4; Page 32-37; 46pp; French.

The present sequence encodes a plant SGS2 polypeptide. SGS2 is an RNA-dependent RNA polymerase and is involved in transgene silencing. Inactivation of SGS2 is used to increase transgene stability and expression in plants, particularly crop plants, especially maize, corn, barley, sorghum, soya, sugar cane, beet, tobacco or cotton plants. Overexpression of SGS2 can be used to increase resistance to viral infection in plants.

Sequence 3591 BP; 1022 A; 714 C; 858 G; 997 T; 0 other;

Query Match 8.28; Score 305.8; DB 22; Length 3591;  
Best Local Similarity 51.08; Pred. No. 3.2e-73;  
Matches 1013; Conservative 0; Mismatches 917; Indels 57; Gaps 10;

QY 1263 TAAGAGGGTCTCTAGTACACACCGTCAAGTTTATTTTGTGGTCCAGAGGTTAATGTTT 1322  
DB 1232 TCAGAGACTAGTATATACCCACACAGAGCTATGCTTACCCACAGAGTTGAGTCT 1291  
QY 1323 CCAATCGGGTCTCCGCAATTTATCTGAAGACATAGATAAATTTCTTCGGTTCCTTTG 1382  
DB 1292 CCAACAGGGTACTCAGGAGATACAAAGCTGTGCTGAAGATTTTTCGGGTACTTTCA 1351  
QY 1383 TTGATGAGGATGGGAGAACTGTATTC-----TACAGACTATTACCAAAA 1429  
DB 1352 TGGATGAAGTATGAGACCAATAATTCGAATGTCTCTCTTATTTGTTGCTCCTATTG 1411  
QY 1430 GCAAGTACTGGAAGTGGT-----CAGGACAAACATATAGAGAGTCTTAT 1478  
DB 1412 TGAAGGATTGACATCAAGTCTTTCTCCAGAAAGCTACGTTTTTAAAGAGTGAAGA 1471  
QY 1479 CAACCTCICGGAAAGCTTTGTAATTTGGTGATAAAATTTGAATTTCTTGCATTTTCAT 1538  
DB 1472 GCATATTAACCGATGGGTTTAACTATGTGTAGAAATACAGTTTTCTTAGCATCTCAG 1531  
QY 1539 CGAGCCAGTTGCGGATAATTCAGTGTGGATGTTTGCATCAAGACCTGGCCTTACTGCAA 1598  
DB 1532 CCAATCAACTGAGACCGCTCTGCATGGTCTTTGCTGAAGACGGGAAACACGTGT 1591  
QY 1599 ATGATATAGAGCTTGGATGGTGTATTTCCGAGATCAAGAAATGTCGCAAAATATGCTG 1658  
DB 1592 CAGATATAAAACATGSGATGGGAAGTTCAAAGA---CAAGAAATGGCAAAATGTGCTG 1648  
QY 1659 CCAGACTTGGTCAATCTTTTGGTCTCTCCAGAGACATTTGAGTGTCTTTAGGCAATGAGA 1718  
DB 1649 CTAGATGGGCTGTGCTTCTCTCCACATATGCCACTGTAGATGTCATGCTCACGAGG 1708  
QY 1719 TTGAAGTATATCCCATGTAAGGTTTCATGGAACAGCTATGCTTTCTGATGGAATG 1778  
DB 1709 TTGA---CACTGAGTTCAGATATAGAGAAATGGGTATGTTTCTCTCAGCGAATG 1765  
QY 1779 GTAAATATCTGGTACATTTGCTATAGATGTGCTCAAAA-----TGTGGCCTTCAAT 1832  
DB 1766 GTACATCACACTCACCTCGCTGACGAAGTAAATGGGAAACTTAAGTTGATGCACT 1825  
QY 1833 ATACCCATCTCTTCCAGATCTGTTATGGTGATATAAGGTTGTGGGTGTGATC 1892  
DB 1826 ACAGCCCTTGTCTTATCAGATACGTTACGCAAGTTTCAAAGGGTGTGCTGCTGGC 1885  
QY 1893 CGGATTCATC-----AATGAAGTTGTCTTTGAGAAAGACATGTCGAAATGAATCAG 1946  
DB 1886 CATCAAAAGTATGGAATCAGGCTAGCCCTCGAGACAGTATGAAGAGTTCTTTCCA 1945  
QY 1947 ACAACATAAGTTAGATGCTCTTGGATGGAGCAATATCAGCCTTGTATCTTAATGCTC 2006  
DB 1946 AACATACGATCTTGGAGATCTGTTCTGAGAGGTTTCAACCTGGTCTTAAATCGC 2005  
QY 2007 AACTGATACGCTCTGCTACACTGGAGTGAAGATGAAGTTCTCCAGCAGAGCAAA 2066  
DB 2006 AGATAATTAACCTCTATCCCGTACTAGGTGTTCCGGATGAATAATCTCGGATATGCGG 2065  
QY 2067 AGGAAGCTAGATCAGCTTGTATCTTCTGATCATCTTTTGAAGCCACAGGAGGCTT 2126  
DB 2066 AATCCATGCTCTATAAAGTGAACCGCATCTTGTATGATACAGATGTGGCATTTGAAGTTC 2125  
QY 2127 TGGATATGATCTCTCTGGAGAGACACTAATATCTCAAGCAATGCTAACTGTGGTT 2186  
DB 2126 T---CACGGCATATGTGCTGAACAGGGGAAACACTGCACTATCATGCTTATGTAGTCAGGTT 2182  
QY 2187 ATAAGCCTGATGTAGCCCTTCTTTTCAATGATGTTGCAAAACCTTCCGCGCATCCAAGT 2246  
DB 2183 TCNAACCAAAACCGCGCATCTACGGGGATGTTGCTTCAGTCAGAAATGACAAAC 2242

QY 2247 TGTCTGATTTGGGACTAGATCAAGAATATTTATTTCCAAATGAAGAACAATGATGGGAT 2306  
DB 2243 TCTGGGCTCTCAGAGAAAATCTCGTATTTTGTGTACTCAGGAAGTGCTTAATGGGTT 2302  
QY 2307 GTTTGGATGAATCCAGAAACCTTTGGAATATGGTCAAGTGTGTTTGTTCAGTTTACTGTGCTG 2366  
DB 2303 GCCTAGACGAAGCAGGATACATTGAACATGGCAATGCTTTTATTAAGTCTCTAAACC-- 2360  
QY 2367 GACATGGAGAGTTTCTGACGATTTACATCCATTTAATAACAGCAGATCCACCAACAGTA 2426  
DB 2361 -GTCTATAGAAAATTTGTTCTCCAAACATGGTCTCGTTTAAAGAGACAAGACAGATC 2419  
QY 2427 ATTTCATTTCTGAAGGAAATGTGTTTTCGAAAAATCCATGCTTTCATCTCTGTTGAT 2486  
DB 2420 TGGAACTAGTTAAAGCTATGTAGCCATTTGAAGAATCCTTCTTCCACCCAGGGGATG 2479  
QY 2487 TTCGTTTTTAAAGCGTGTAAATTTTCGAGCGCTGCACACATGGTAGATTTGTTGTAT 2546  
DB 2480 TAAAGGATTTAGAGCTGTTGATGTACCCAGCTGCTCACAATGTATGACTGCCATTATT 2539  
QY 2547 TCCTCAGAAAGGAAAGAACCTCATCCGAATGAATGTTCTGGGAGTGAATTTGGATGGG 2606  
DB 2540 TCCCTCAGAAAGGTGATAGCCGCATACAAACGAAGCTTCTGGCACTGACCTTGACGGG 2599  
QY 2607 ATATCTACTTTTGTCTGGGATCAAGACATGATCCCGC-----AAGGCAAGTCCAGC 2660  
DB 2600 ACCTGTACTTTTGGTGGGATCAGAAATCATCCTCCCAACAGAAAGCTATTCGG 2659  
QY 2661 CGATGAATATCCTCCAGACCCAGCATACAGTTGACCATGATGTCACAAATTTGAGAAAG 2720  
DB 2660 CCATGATATGATGAGCTGAAGAGAGAGTTTAGCCGCTGCTGTCACACCACGAGCA 2719  
QY 2721 TTGAAGAGTACTTCCACAACTATATTGTGAATCAGAGTTTGGAAATCATAGCAATGCC 2780  
DB 2720 TAATCGATTTCTTTGCAAGAACTTGGCAATGAGCAGTTGGGCAAAATTTGCAATGCC 2779  
QY 2781 ATCTGTTTTCAGACAGAGAACTGATATGGCCATGATGATCCATCCAAAACTTTG 2840  
DB 2780 AGTCTGTTCACTGATAGAGTGTAGTGGACCATGGACGAAGATGTTTCTACTG 2839  
QY 2841 CTGAGCTCTTTCAATTCAGTGGACTTTCCAAAGACTGCTGTTCCCGCTGAATACCAT 2900  
DB 2840 CAGAACTAGCTGCCACTGCAGTTGATTTCCCAAGACAGGGGAAATTTGTGTCATGCCCT 2899  
QY 2901 CTGAGTTGCGCCTTAAAGAATACCCAGCTTCATGGATAAGCCGACAGACAGCAGCTATA 2960  
DB 2900 TCCACCTTAAACCAAACTCTACCCAGTTTCATGGAAAGAGACTACCAAACTTACA 2959  
QY 2961 TCTCAGAAAGAGTTATTTGAAAGCTTTTCAGGAAAGTGA---AGGACAAAGCAGCTCAGG 3017  
DB 2960 AGTCGAAACAAATCTTTGGTGGCTTTTACAGACGGGTAAAGAGGTTTATGATGAAGATG 3019  
QY 3018 CTAGCTCTATCGCGACCTTCCAGAGAGATGTTCAAGGAGATCATATGATGCTGATATGG 3077  
DB 3020 CAGAGCTTCTCTCAGAAAGACAGACCCCAAGTCCCATCCCTTATGAGCTGTTCTTG 3079  
QY 3078 AAGTTGATGATTTGAAGATTACATTGACGAAGCTTTTGTACTACAAACTGAATATGACA 3137  
DB 3080 AAATACCGGATTTGAAGATTGATCCCTGAGGATGGGCTCACAATGTTTGTACGACG 3139  
QY 3138 ACAAGCTGGGTAATTTAATGACTACTATGGCATAAAACAGAGGCTGAAATACTTAGTG 3197  
DB 3140 GGCACCTATGGTCTTTCTTGGGCAATCAAGGTGCAAGAGAGAGAGATTTGTGACGG 3199  
QY 3198 GTGGCAT 3204  
DB 3200 GTCACAT 3206

RESULT 8  
AAC63741  
ID AAC63741 standard; cdna; 2816 BP.  
XX









Db 836 GAAGCTGATATGACCCCTGGATTGATGCTAGGGGCTCTATTGATTGTTGTTTCT 895  
QY 2552 CAGAAAGGAAAGAACCTCATCCGAATGAATGTTCTGGAGTGTATTTGGATGGGATATC 2611  
Db 896 CAGAGAGGGGAAAGCCCTCATCCGAATGAATGCTCCGGGGCGATTGATGGGACCTC 955  
QY 2612 TACTTTGTTGCTGGATCAAGACATGATCCCGCAAGCGAAGTCCAGCCGATGGAATAT 2671  
Db 956 TCTTTTATTACTTGGATGACAACTGATTCCGGAGAAGTTGATGCCACCTATGACTATC 1015  
QY 2672 COTCCAGCACCAGCATACAGTTGACCATGATGTCACAAATGAGGAAGTTTGAAGATAC 2731  
Db 1016 ACTGCAAGAGGCCACGCATATGACCATGCTGTACCTTGAGGAATTCAGAAGCAC 1075  
QY 2732 TTCACCACTATATTTGAATGACAGTTTGGGAATCATAGCAATGCCCATGTCGTATT 2791  
Db 1076 TTCGTCAGTTACATGATAAAGATACCTCGGTGCTCCAGCGCCACTTGATCCAC 1135  
QY 2792 GCAGACAGAACCTGATATGCCCATGAGTGCAGCCCGAGTCCAGCTGGCGCTCGCAC 1195  
Db 1136 GCAGACCTGATCCCTGAAAGCTCGAGCCCGAGTCCAGCTGGCGCTCGCAC 1195  
QY 2852 TCAATTGACGTGAGTATCCAAAGACTGGTGTTCCTCCGCTGAAATACC 2898  
Db 1196 TCATGGGGTGCAGTTCGCAAGAGCGGAGCTCAGCCCAAGATTCC 1242

## RESULT 10

AAC63743

ID AAC63743 standard; cDNA; 740 BP.

AC AAC63743;

DT 07-FEB-2001 (first entry)

DE Maize RNA-directed RNA polymerase EST contig, SEQ ID NO: 13.

KW Maize; plant; RNA-directed RNA polymerase; expressed sequence tag; EST;

KW gene mapping; gene marker; plant virus resistance; plant breeding; ss.

XX Zea mays.

XX WO200060097-A1.

XX 12-OCT-2000.

XX 06-APR-2000; 2000WO-US09105.

XX 07-APR-1999; 99US-0128094.

XX (DUPO) DU PONT DE NEMOURS &amp; CO E I.

XX (PION-) PIONEER HI-BRED INT INC.

XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;

XX WPI; 2000-679376/66.

XX P-PSDB; ABE28534.

XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful

XX for controlling gene expression and providing mechanisms to engineer

XX plant virus resistance.

XX Disclosure; Page 55; 62pp; English.

XX The present sequence is one of a number of cDNA molecules which encode

XX plant RNA-directed RNA polymerase proteins. The sequences are useful as

XX probes for genetically and physically mapping genes, and as markers for

XX traits linked to those genes. They are useful for controlling gene

XX expression and provide mechanisms to engineer plant virus resistance.

XX They are also useful for plant breeding to develop lines with desired

XX phenotypes.

SQ Sequence 740 BP; 217 A; 141 C; 205 G; 166 T; 11 other;

Query Match 7.4%; Score 276.2; DB 21; Length 740;

Best Local Similarity 64.2%; Pred. No. 2e-65;

Matches 451; Conservative 0; Mismatches 231; Indels 21; Gaps 2;

QY 2771 GCAAAATCCCATCGTATTGTCAGACAGAGAACCTCATATGGCCCATGAGTGCATGCG 2830

Db 2 GCAATCCGATGTTGTTTTCAGATCAGAACGATGATGAAGCTGAGAGTCCACCGTGC 61

QY 2831 AAAAACTTCTGAGCTCTTTTCAATGTCAGTGGACTTCCAAAGACTGTGTTCGGCT 2890

Db 62 GTTCAACTGGCCAGCTCTTCTATAGCTGTGATTTCCCAAGACTGAGTGCCTGCT 121

QY 2891 GAATATCCATCTCAGTTGCGCCCTTAAAGATACCCAGACTTCATGGATAAGCGGACAAG 2950

Db 122 CTGATTCACATGAGCTACATGTCAGGAGTATCTCTGACTTCATGGAGAATTCGACAAA 181

QY 2951 ACCAGCTTATATCTCAGAAAGAGTTATTGAAAGCTTTTCAGGAAAGTGAAGGACAAAGCA 3010

Db 182 GTCACCTATGAATCAAAGGGTGTGTCGGGAAGCTCTATAGGAAATATAAAGAGCACACA 241

QY 3011 CCTCAGCTAGCTCTATTCGGGACCTTCACAAGAGATGTTGCAAGGAGATCATATGATGCT 3070

Db 242 CCACA-----CATAAAGCACTTCACGAGGAAGTGGCAAGCGGCTTTATGACACC 292

QY 3071 GATATGGAAGTTGATGATTGAAAGATTACATTGACCAAGCTTTTGACTACAAAAGTGA 3130

Db 293 GATTGATTGTTGATGCTATGAAGATTACATTACTGAGGCTATAGAGTTCAAGGAAGAG 352

QY 3131 TATGACAAAGCTGGGTAAATTTAATGACTACTATGGGCATATGAGGCAATAAAGAGAGCTGAATA 3190

Db 353 TACGATTTTCAGGCTGGGTAAATCTTATGACCACTATGCGCATATAAAGAGTGAAGCTGAGATA 412

QY 3191 CTTAGTGTGTCATTATGAAAGCATCAAAAACCTTTGACCGCAGAAAAGATGCTGAGGCC 3250

Db 413 ATAACTGGATGATTCTAAAGATGGCAAGATTTTCCACAGAGTAGNATGCTGATGCA 472

QY 3251 ATTAGTGTGTCAGGGCTTGAAGAGAGGCAAGAGCCTGGTTCAGAGGCCGTA-- 3308

Db 473 ATTAGAATGGCGNGAGATCTTTGAGGAAAGAGCTAGTTCGNGGNTCAATGAGATGAGC 532

QY 3309 -----ATGATATAGATGACATGTTACCAAAGCTTCGGCTTGGTACCAGCTTACA 3358

Db 533 ACAGGAGAGGATGGCCAAAGATGCCATGGAGGCCAAGGCTCTCTTGGNACCATGTTACT 592

QY 3359 TATCATCTACATATGGGTGCTACATCAGGGGTGAAAGAGCTCATTTTCATTAGC 3418

Db 593 TATCATCAGCAGTACTGGGCGAGCTACAATGAAGGTATGATCGNGCGCATCTTATTAGC 652

QY 3419 TTTCCCTGGTGTGTTTATGACCACTAATCCAGATTAAAGAGG 3461

Db 653 TTCCATGGNGCGGATATGACAAGCTTGGGGGCGATCAAGCANG 695

## RESULT 11

AAH77703

ID AAH77703 standard; DNA; 6863 BP.

AC AAH77703;

XX 13-NOV-2001 (first entry)

XX Nucleotide sequence of the SGS2 gene.

XX SGS2; RNA-dependent RNA polymerase; transgene silencing;

XX transgene stability; crop plant; viral resistance; ss.

XX Arabidopsis thaliana.

XX Key Location/Qualifiers

XX Promoter 1..850

XX /\*tag= a

```

FT exon      851..3564
FT           /*tag= b
FT           /number= "1"
FT intron    3565..3986
FT           /*tag= c
FT           /number= "1"
FT exon      3987..4862
FT           /*tag= d
FT           /number= "2"
FT terminator 4863..6863
FT           /*tag= e
XX
XX FR2804128-A1.
XX
XX PD 27-JUL-2001.
XX
XX PF 26-JAN-2000; 2000FR-0001007.
XX
XX PR 26-JAN-2000; 2000FR-0001007.
XX
XX (RHOB-) RHOBIO.
XX (INRG ) INST NAT RECH AGRONOMIQUE.
XX
XX Beclin C, Elmayer T, Mourrain P, Vaucheret H;
XX WPI; 2001-543303/61.
XX
XX A new plant SGS2 gene involved in encoding an RNA-dependent RNA
XX polymerase and in transgene silencing, increases transgene stability
XX and expression in transgenic plants when it is inactivated -
XX
XX Claim 2; Page 30-32; 45pp; French.
XX
XX The present sequence encodes a plant SGS2 polypeptide. SGS2 is an
XX RNA-dependent RNA polymerase and is involved in transgene silencing.
XX Inactivation of SGS2 is used to increase transgene stability and
XX expression in plants, particularly crop plants, especially maize, corn,
XX barley, sorghum, soy, sugar cane, beet, tobacco or cotton plants.
XX Overexpression of SGS2 can be used to increase resistance to viral
XX infection in plants.
XX
XX Sequence 6863 BP; 1862 A; 1354 C; 1485 G; 2162 T; 0 other;
XX
XX Query Match      5.9%; Score 220.4; DB 22; Length 6863;
XX Best Local Similarity 51.5%; Pred. No. 1.4e-49;
XX Matches 733; Conservative 0; Mismatches 641; Indels 48; Gaps 8;
XX
XX 1263 TAAGAGGGTCTAGTACACCATGCAAGTTATTTTGGTCCAGAGGTTAATGTT 1322
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 2081 TCAGAAGACTAGTAATACCCCAACAGAGGCTATGCTGCTACCCCAAGTGGAGCTCT 2140
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 1323 CCAATCGGGTCTCGGCAATTATCTGAAGACATAGATAAATTTCTGCTGTTCTTTG 1382
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 2141 CCAACAGGGTACTCAGGAGATACAAGCTGTGCTGAAGATTTTGGGGTAACTTCA 2200
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 1383 TTGATGAGAGTGGGAGAACTGATTC-----TACAGCTTATTACCAAAA 1429
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 2201 TGGATGAAGTATGAGACCAATAATGAAATGTTCTCTTACTTCTGCTCTATG 2260
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 1430 GCAAGTACTGGAAGTGGT-----CAGGACAAACATCTATGAGAGGATCTTAT 1478
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 2261 TGAAGGATTTGACATCAAGTTCTTCTCCAGAGACCTACGTTTAAAGAGTGAAGA 2320
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 1479 CAACTCTCGGAAAGCTTTGTAATGGTGATATAAAATTTGAATTTCTTCATTTTCAT 1538
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 2321 GCATATTACCGATGGGTTAACTATGTGTAGAAAATACAGTTTCTAGCATCTCAG 2380
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 1539 CGAGCCAGTTGGGGATAATTCAGTGTGGATGTTTGCATCAAGACCTGGCCCTTACGAA 1598
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 2381 CCAATCACTGAGACCGCTCGATGGTCTTGTGTAACCGGAAACACGTGTGT 2440
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 1599 ATGATATAAGACTGGATGGGTGATTTTTCGAGATCAAGAAATGTCGCAAAATATGCTG 1658
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 12  
ABL71773

```

Db 2441 CAGATATAAAACATGGATGGGGAAGTT---CAAAGACAAAGAAATGTGGCAAAATGTGCTG 2497
QY CCAGACTTGGTCAATCTTTTGGTTCCTCCAGAGACTTTGAGTGTCTTAGCCATGAGA 1718
Db 2498 CTAGGATGGCCCTGTCTCTCTCCACATATGCCACTGTAGATGTCATGCCCTCAGAGG 2557
QY TTGAAGTTATCCCGATGTAAAGTTTCATGGAAACAGCTATGTCTTTCTTGATGGAATTG 1778
Db 2558 TTGA---CACTGAGGTTCCAGATATTGAGAGAAATGGGTATGTTTCTCTGACGGAATTG 2614
QY GTAAATATCTGCTGACTTTGCTCATAGAGTTGCCCTCAAAA-----TGTGGCTTCAAT 1832
Db 2615 GTAAATATCTGCTGACTTTGCTCATAGAGTTGCCCTCAAAA-----TGTGGCTTCAAT 2674
QY ATACCCCATCTGCTTCCAGATTCGTTATGGTGATATAAAGGTGTGTGGGTGTGATC 1892
Db 2675 ACAGCCCTTGTGCTTATCAGATACGTTAGCAGGTTTCAAAGGGGTGTGTGCTCGTTGCC 2734
QY C-----GGATTTCATCAATGAAGTTGCTTTGAGAAGAGCATGTCGAAATATGAATCAG 1946
Db 2735 CATCAAAAAGTGATGGAATCAGGCTAGCCCTTCGACACAGTATGAAGAGTCTTTTCCA 2794
QY ACAACATAAAGTTAGATGTCCTTGGATGGAGCAAAATATCAGCCTTGTATCTTAATCTGTC 2006
Db 2795 AACATACGATCTTGGAGATCTGTTCTCGACGAGGTTTCAACCTGGTCTTAAATCGGC 2854
QY AACTGATTACGCTCTGCTCTACACTTGGAGTGAAGTGAAGTCTCTCGAACAGCAAA 2066
Db 2855 AGATAATTACCTTCATCCGTTAGGTGTTCCGGATGAATATTTCTGGATATGACAG 2914
QY AGGAAGCTGTAGTACGCTTGTCTATCTTGATGATGATTTTGAAGGACAGAGGCTT 2126
Db 2915 AATCCATGCTCTATAAATCAACCGCATCTTGATGATACAGATGTGGCATTTGAAGTTC 2974
QY TGGAAATGTGCTCTCGGAGAGAACTAATATTCTCAGGCAATGCTTAACCTGTTG 2186
Db 2975 TCACGGCATCATGTGCTGAACAGGAACTGCACTGCACTAT---CATGCTTAGTGCAGTT 3031
QY ATAAGCCTGTATGCTGAGCCCTTCTTTTCAATGATGTGTTGAAACCTTCCGCGCATCAAAGT 2246
Db 3032 TCAACCAAAAACCGAGCGCATCTACGGGGATGTTGTTCTTCAGTCAGAATTCACAAC 3091
QY TGCTCGATTTGGGACTAGATCAAGATATTTTATTCAAATGGAAGAACAAATCATGGAT 2306
Db 3092 TCTGGGCTCTCAGAGAAAATCTCGTATTTTGTACTTCAGGAAGGTGGCTAATGGTT 3151
QY GTTTGGATGAATCCAGAACCTTGGATATGCTCAGGTGTTTGTTCAGTTTACTTGTGCTG 2366
Db 3152 GCCTAGACGAGCAGGATACCTTGAACATGSCCAATGCTTATTCAAGTCTCTAAACCGT 3211
QY GACATGAGAGTCTTCTGAGATTTACATCCATTTAATACACAGATCCACCACAGTA 2426
Db 3212 CTATAGAAAATTTGTTCTCCAAACATGGGTCTCGTTTAAAGGAGACAAAGACAGATCTGG 3271
QY ATTTCTATCTGAAGGAAATGTGTTGTTGCAAAAATCCATCTTCATCTCTGTTGATA 2486
Db 3272 AAGTAGTT---AAGGCTATGAGCCATGCTTAAGATCCTTCTTCAACCCAGGGATG 3328
QY TTCGTGTTTAAAGGTGTAATATGTTTCGACGCTGCAACACATGTTAGTGTGTTGAT 2546
Db 3329 TAAGGATTTTAGAGCTGTTGATGTATGCCAGCTGCATCATCATGATGTGCTTATTT 3388
QY TCCTCAGAAAGGAAAGAACCTCATCCGAATGAATGTTCTCTGGAGTGAATTTGGATGGGG 2606
Db 3389 TCCTCAGAAAGGTGATAGCCCGCATACAACGAGCTTCTGCGAGTGACCTTGACGGG 3448
QY ATATCTACTTTTGTGCTGGATCAAGACATGATCCCGCAA 2648
Db 3449 ACCTGTACTTTGTGGCTTGGGATCAGAAACTCATCCCTCCA 3490

```

QY	2560	ARAAAGACCTCATCCGAATGAATCTCTGGAGTATTTCGATGGCGATATCTACTTGT	2619
Db	134	TGAGAGCCGCGACGGAATGAACCATCTGGGAGTGATCTTGATGGGATCTATACTCGT	193
QY	2620	TTGCTGGGATCAAGACATGATCCCGCCAAAGCAA	2653
Db	194	AACATGGGATGAAACCTTATACCACTGGTAA	227
<p>RESULT 13</p> <p>ABL71722</p> <p>ID ABL71722 standard; cDNA; 274 BP.</p> <p>XX ABL71722;</p> <p>XX AC</p> <p>XX DT</p> <p>XX 14-MAY-2002 (first entry)</p> <p>XX DE</p> <p>XX Corn tassal-derived polynucleotide (cdps) SEQ ID NO:1096.</p> <p>XX DE</p> <p>XX Corn; corn tassal-derived polynucleotide; cdps; hybrid breeding; CDPs;</p> <p>KW inheritance; characteristic; growth; development; disease resistance;</p> <p>KW environmental adaptability; quality; yield; molecular marker;</p> <p>KW multigene trait; plant breeding; corn tassal; gene; ss.</p> <p>XX OS</p> <p>XX Zea mays.</p> <p>XX US2001051335-A1.</p> <p>XX PN</p> <p>XX 13-DEC-2001.</p> <p>XX PD</p> <p>XX 16-APR-1999; 99US-0294093.</p> <p>XX PF</p> <p>XX 21-APR-1998; 98US-082567P.</p> <p>XX PR</p> <p>XX {LALG/} LALGUDI R V.</p> <p>PA {ITOL/} ITO L Y.</p> <p>PA {SHER/} SHERMAN B K.</p> <p>PA PA</p> <p>PI Lalgudi RV, Ito LY, Sherman BK;</p> <p>PI WPI; 2002-163647/21.</p> <p>DR</p> <p>XX</p> <p>PT Novel purified corn tassal-derived polynucleotide useful for</p> <p>PT determining altered gene expression, to recover regulatory elements and</p> <p>PT to follow inheritance of desirable characteristics through hybrid</p> <p>PT breeding programs -</p> <p>XX</p> <p>XX Claim 1; SEQ ID 1096; 201pp; English.</p> <p>XX</p> <p>XX The present sequence describes a purified corn tassal-derived</p> <p>CC polynucleotide sequence (cdps) comprising a nucleic acid sequence</p> <p>CC selected from those given in ABL70537 to ABL76833. The cdps sequences</p> <p>CC encode corn tassal-derived polypeptides (CDPs). The cdps sequences (I)</p> <p>CC can be used for determining altered gene expression, to recover</p> <p>CC regulatory elements and to follow inheritance of desirable</p> <p>CC characteristics through hybrid breeding programs. (I) are also useful</p> <p>CC in the evaluation, and alteration of desired characteristics associated</p> <p>CC with growth and development, disease resistance, environmental</p> <p>CC adaptability, quality and yield, and as molecular markers for studying</p> <p>CC inheritance of multigene traits in a plant breeding program. (I) can be</p> <p>CC used to produce a tassal-specific profile of gene transcription, a</p> <p>CC transcript image, to clone regulatory elements for use in transformatio</p> <p>CC vectors, to express a polypeptide, to identify, isolate or extend</p> <p>CC identical or related corn tassal nucleic acid sequences from DNA</p> <p>CC libraries, in nucleic acid hybridisation or amplification technologies</p> <p>CC as query sequences to determine homology of known sequences, as probe</p> <p>CC for use in Southern or Northern hybridisation, and to identify the</p> <p>CC presence of and/or to determine the degree of similarity between two</p> <p>CC (or more) nucleic acid sequences.</p> <p>XX</p> <p>XX Sequence 274 BP; 68 A; 58 C; 72 G; 76 T; 0 other;</p> <p>XX SQ</p>			
QY	Query Match		2.6%; Score 95.4; DB 24; Length274;





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 19:22:12 : Search time 158 seconds  
(without alignments)  
7867.093 Million cell updates/sec

Title: US-09-782-874-1  
Perfect score: 3731  
Sequence: 1 GAAATATCTTACTTACTT.....AGTTTCATCTTCTCTCAA 3731

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications, NA:\*

- 1: /cgn2\_6/ptodata/1/pubna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubna/PC1\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubna/PC1US\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3731	100.0	3731	10	US-09-782-874-1
2	106.8	2.9	282	10	US-09-294-093B-1147
3	95.4	2.6	274	10	US-09-294-093B-1096
4	75	2.0	260	10	US-09-878-574-10527
5	40.4	1.1	23071	10	US-09-764-864-1673
6	40.2	1.1	1904	10	US-09-864-761-10600
7	39.2	1.1	2073	10	US-09-815-242-8362
8	39.2	1.1	2109	10	US-09-815-242-4555
9	38	1.0	5191	10	US-09-962-832-124
10	37.8	1.0	2720	10	US-09-847-665-3
11	37.8	1.0	3699	10	US-09-847-665-6
12	37.6	1.0	567	10	US-09-974-300-1446
13	36.4	1.0	499	10	US-09-864-761-412
14	36.4	1.0	3191	10	US-09-939-980-194
15	36.2	1.0	1196	10	US-09-764-847-1401
16	36	1.0	273	10	US-09-923-876-6245
17	35.6	1.0	2397	10	US-09-909-320-324
18	35.6	1.0	2397	10	US-09-909-088B-324
19	35.6	1.0	2397	12	US-10-052-586-29

C	20	35.4	0.9	53226	10	US-09-818-264-3	Sequence 3, Appli
	21	35	0.9	1275	10	US-09-771-838A-4	Sequence 4, Appli
	22	35	0.9	1458	10	US-09-771-838A-19	Sequence 17, Appl
	23	35	0.9	1796	10	US-09-771-838A-17	Sequence 19, Appl
	24	35	0.9	1845	10	US-09-771-838A-5	Sequence 5, Appli
C	25	35	0.9	2000	10	US-09-887-575-126	Sequence 126, App
	26	35	0.9	2046	10	US-09-771-838A-6	Sequence 6, Appli
	27	35	0.9	3023	10	US-09-079-892-4	Sequence 4, Appli
	28	35	0.9	3048	12	US-10-044-090-685	Sequence 685, App
	29	35	0.9	111282	12	US-10-094-989-3	Sequence 3, Appli
	30	35	0.9	397658	10	US-09-813-320-3	Sequence 3, Appli
	31	34.6	0.9	471	10	US-09-815-242-4575	Sequence 4575, Ap
	32	34.6	0.9	549	10	US-09-815-242-8154	Sequence 8154, Ap
C	33	34.6	0.9	1134	10	US-09-815-242-7451	Sequence 7451, Ap
	34	34.6	0.9	1883	10	US-09-796-256A-3	Sequence 3, Appli
	35	34.6	0.9	4756	10	US-09-982-091A-3	Sequence 3, Appli
	36	34.6	0.9	32146	10	US-09-764-860-797	Sequence 797, App
C	37	34.6	0.9	32248	10	US-09-764-860-802	Sequence 802, App
	38	34.4	0.9	304	10	US-09-294-093B-623	Sequence 623, App
C	39	34.4	0.9	329	10	US-09-960-352-7689	Sequence 7689, Ap
	40	34.4	0.9	501	10	US-09-864-761-12414	Sequence 12414, A
	41	34.4	0.9	1537	10	US-09-740-273-1	Sequence 1, Appli
C	42	34.4	0.9	8188	10	US-09-070-927A-21	Sequence 21, Appli
	43	34.2	0.9	3831	12	US-10-051-952-5	Sequence 5, Appli
C	44	34	0.9	42999	10	US-09-740-028-3	Sequence 3, Appli
C	45	33.8	0.9	1867	10	US-09-846-590B-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1  
US-09-782-874-1  
; Sequence 1, Application US/09782874  
; Patent No. US20010023067A1  
; GENERAL INFORMATION:

APPLICANT: Wassenegeger, Michael  
Riedel, Leonhard  
Schiebel, Winfried  
Sanger, Heinz

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING

POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN  
RNA-DIRECTED RNA POLYMERASE (RRRP)

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/09/782,874

APPLICATION NUMBER: US/09/782,874

FILING DATE: 08-Feb-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/811,583

FILING DATE: 05-MAR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Haley, James F.

REGISTRATION NUMBER: 27,794

REFERENCE/DOCKET NUMBER: MPG-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-596-9000

TELEFAX: 212-596-9090

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3731 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Tomato  
FEATURE:

US-09-782-874-1  
SEQUEN

```
Query Match 100.0%; Score 3731; DB 10; Length 3731;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	1	GAATATCTTTACTTACTTCCACCGGATTCATCACTCCCTCAAGTCCTTTGTGT	60
Db	1	GAATATCTTTACTTACTTCCACCGGATTCATCACTCCCTCAAGTCCTTTGTGT	60
QY	61	GTGTGATATAAATTTGGTTGTCTCAGTTTCAGTCACTACTGCTGGGTAGTTTTAT	120
Db	61	GTGTGATATAAATTTGGTTGTCTCAGTTTCAGTCACTACTGCTGGGTAGTTTTAT	120
QY	121	TTTGCATAACTCCAGGGGTATTCAGTTGGTGTAGCATTTGAAAGTCGAAGTCGAACCTGCACCT	180
Db	121	TTTGCATAACTCCAGGGGTATTCAGTTGGTGTAGCATTTGAAAGTCGAAGTCGAACCTGCACCT	180
QY	181	GGAAATTGGCTACATGGGAAGACAATTCAGGTTTCGGATTCCCTTATCTTCTCTCTGC	240
Db	181	GGAAATTGGCTACATGGGAAGACAATTCAGGTTTCGGATTCCCTTATCTTCTCTCTGC	240
QY	241	GGAAAGTGGTTAAGTCATCTCTACAGAAAATATACAGGATATGGAACGTATGTGTCATTGGA	300
Db	241	GGAAAGTGGTTAAGTCATCTCTACAGAAAATATACAGGATATGGAACGTATGTGTCATTGGA	300
QY	301	GGTTAAACAGTCCAAAGGAGGATCTAGAGCATTTGCGAAAGTTCAATTTGCCGACAACAT	360
Db	301	GGTTAAACAGTCCAAAGGAGGATCTAGAGCATTTGCGAAAGTTCAATTTGCCGACAACAT	360
QY	361	AAGTGCTGACAAAATCATCACTTTGGCTAATAACAGGCTGATATTTGGGCTCTTCTTATTT	420
Db	361	AAGTGCTGACAAAATCATCACTTTGGCTAATAACAGGCTGATATTTGGGCTCTTCTTATTT	420
QY	421	GAAGGCTTGGGAAATGAAACTGATATTTGTCGAACCTGGGGCATATGTGGTCACAGTGA	480
Db	421	GAAGGCTTGGGAAATGAAACTGATATTTGTCGAACCTGGGGCATATGTGGTCACAGTGA	480
QY	481	TGGCATAACTTTGAATTTGCGATGTCAGATATCAGATGACAAGTTTGCAGTGTGGGAAG	540
Db	481	TGGCATAACTTTGAATTTGCGATGTCAGATATCAGATGACAAGTTTGCAGTGTGGGAAG	540
QY	541	TACAGAAGTTTCAATTCAAATTTGGCATTTGGCATTTGGATTAAGAAATTTTTTTCTTTTTATCTAG	600
Db	541	TACAGAAGTTTCAATTCAAATTTGGCATTTGGCATTTGGATTAAGAAATTTTTTTCTTTTTATCTAG	600
QY	601	TGGTTACGCTGACTATAAAGCTTCAGCTTTCATATGAAATATATGCGAGGTTGTGCTCCA	660
Db	601	TGGTTACGCTGACTATAAAGCTTCAGCTTTCATATGAAATATATGCGAGGTTGTGCTCCA	660
QY	661	TCGTCCATATGGTCAAAATGCTCAGTTTCTCTCATACAGTATTTTGGTGCTCCCTCGGAT	720
Db	661	TCGTCCATATGGTCAAAATGCTCAGTTTCTCTCATACAGTATTTTGGTGCTCCCTCGGAT	720
QY	721	CTATAAGAGACTTGAAACCTCCTGTTATAGCTTCTTTAAGGAAACTCCTGTATCAGTG	780
Db	721	CTATAAGAGACTTGAAACCTCCTGTTATAGCTTCTTTAAGGAAACTCCTGTATCAGTG	780
QY	781	GGTGAGGACAACAGATTTCCCTCCATCTTGGATAGGGCTATCTTCTAGCTTATGTTTGA	840
Db	781	GGTGAGGACAACAGATTTCCCTCCATCTTGGATAGGGCTATCTTCTAGCTTATGTTTGA	840
QY	841	GTTCCGTTAGGGGTGTCGTCTTCCAAATTTTCGAGGAAAGTTTTTTTCCACTATGCGAAGC	900



Db 1921 AAAGAGATGTCGAAATATGAATCAGACACACATAAAAGTTAGATGCTCCTGGATGGAGCAA 1980  
QY 1981 ATATCAGCCTTGTTATCTTAATCGTCAACTGATTAACGCTCTGTGCTACACTTTGGAGTGA 2040  
Db 1981 ATATCAGCCTTGTTATCTTAATCGTCAACTGATTAACGCTCTGTGCTACACTTTGGAGTGA 2040  
QY 2041 AGATGAAGTTCTCGAACAGAGCAAAAGGAGCTGTAGATCAGCTTGATCTATCTTGCA 2100  
Db 2041 AGATGAAGTTCTCGAACAGAGCAAAAGGAGCTGTAGATCAGCTTGATCTATCTTGCA 2100  
QY 2101 TGATCTTTTGAAGCACAGGAGGCTTTGGAATTGATCTCTCGGAGAGAACACTAATAT 2160  
Db 2101 TGATCTTTTGAAGCACAGGAGGCTTTGGAATTGATCTCTCGGAGAGAACACTAATAT 2160  
QY 2161 TCTCAAGGCAATGCTAAACTGTGTTATAGCCTGATGCTGAGCCCTTTCTTCAATGAT 2220  
Db 2161 TCTCAAGGCAATGCTAAACTGTGTTATAGCCTGATGCTGAGCCCTTTCTTCAATGAT 2220  
QY 2221 GTTCCAAACCTTCGCGCATCCAAAGTTGCTCGATTTCGCGACTAGATCAAGAATATTTAT 2280  
Db 2221 GTTCCAAACCTTCGCGCATCCAAAGTTGCTCGATTTCGCGACTAGATCAAGAATATTTAT 2280  
QY 2281 TCCAAATGGAGACATGATGGATGTTGGATGAATCCAGAACCTTGGAAATATGGTCA 2340  
Db 2281 TCCAAATGGAGACATGATGGATGTTGGATGAATCCAGAACCTTGGAAATATGGTCA 2340  
QY 2341 GGTGTTTGTTCAGTTTACTGCTGACATGGAGAGTTCCTGAGGATTTACATCCATT 2400  
Db 2341 GGTGTTTGTTCAGTTTACTGCTGACATGGAGAGTTCCTGAGGATTTACATCCATT 2400  
QY 2401 TAATAACAGCAGATCCACCACAGTAATTTCAATTCGAAGGAAATGTGTTGTTGCAAA 2460  
Db 2401 TAATAACAGCAGATCCACCACAGTAATTTCAATTCGAAGGAAATGTGTTGTTGCAAA 2460  
QY 2461 AAATCCATGTTTGCATCCTGCTGATATTCGTTTAAAGGCTGTAATGTTTCGAGCGCT 2520  
Db 2461 AAATCCATGTTTGCATCCTGCTGATATTCGTTTAAAGGCTGTAATGTTTCGAGCGCT 2520  
QY 2521 GCACACATGTTAGATGTTGTTATTCCTCAGAAAGGAAAGACCTCATCCGAATGA 2580  
Db 2521 GCACACATGTTAGATGTTGTTATTCCTCAGAAAGGAAAGACCTCATCCGAATGA 2580  
QY 2581 ATGTTCTGGAGTATGTTGATGGGATATCTACTTGTGTTGCTGGATCAAGACATGAT 2640  
Db 2581 ATGTTCTGGAGTATGTTGATGGGATATCTACTTGTGTTGCTGGATCAAGACATGAT 2640  
QY 2641 CCCGCCAAGCAAGTCCAGCGGATGGAATATCCTCCAGCACCCAGCATACAGTTGGNCCA 2700  
Db 2641 CCCGCCAAGCAAGTCCAGCGGATGGAATATCCTCCAGCACCCAGCATACAGTTGGNCCA 2700  
QY 2701 TGATGTCACAAATGAGGAAGTTGAAGTACTTCACTTGTGTTGCTGGATCAAGACATGAT 2760  
Db 2701 TGATGTCACAAATGAGGAAGTTGAAGTACTTCACTTGTGTTGCTGGATCAAGACATGAT 2760  
QY 2761 GGGATCATAGCAATCCCATGCTCGATTTGAGAGAGAGAACCTGATATGGCCATGAG 2820  
Db 2761 GGGATCATAGCAATCCCATGCTCGATTTGAGAGAGAGAACCTGATATGGCCATGAG 2820  
QY 2821 TGATCCATGCAAAACCTTCTGAGCTCTTTCAATTCGAGTGGACTTTCCAAAGACTGG 2880  
Db 2821 TGATCCATGCAAAACCTTCTGAGCTCTTTCAATTCGAGTGGACTTTCCAAAGACTGG 2880  
QY 2881 TGTTCCCGCTGAATACCATCTCAGTTCGCGCTTAAGATATCCAGACTTCATGATGA 2940  
Db 2881 TGTTCCCGCTGAATACCATCTCAGTTCGCGCTTAAGATATCCAGACTTCATGATGA 2940  
QY 2941 GCCGGACAAGACCAGCTATATCTCAGAAAGAGTATTTGAAAGCTTTTCAGGAAAGTGAA 3000  
Db 2941 GCCGGACAAGACCAGCTATATCTCAGAAAGAGTATTTGAAAGCTTTTCAGGAAAGTGAA 3000  
QY 3001 GGACAAAGACCTCAGGCTAGCTCTATCGCGACCTTCACAAGAGATGTTGCAAGAGATC 3060  
Db 3001 GGACAAAGACCTCAGGCTAGCTCTATCGCGACCTTCACAAGAGATGTTGCAAGAGATC 3060

QY 3061 ATATGATGCTGATATGAAGTTGATGATTGAAGATTACATTGACGAAGCTTTTGACTA 3120  
Db 3061 ATATGATGCTGATATGAAGTTGATGATTGAAGATTACATTGACGAAGCTTTTGACTA 3120  
QY 3121 CAARAACCTGAATATGACAAACAGCTGGTAAATTTAATGGACTACTATGGCATAAACACGA 3180  
Db 3121 CAARAACCTGAATATGACAAACAGCTGGTAAATTTAATGGACTACTATGGCATAAACACGA 3180  
QY 3181 GGTGAAATACCTTAGTGTGGCATTTATGAAGGCATCAAAAACCTTTTGACCGCAGAAAAGA 3240  
Db 3181 GGTGAAATACCTTAGTGTGGCATTTATGAAGGCATCAAAAACCTTTTGACCGCAGAAAAGA 3240  
QY 3241 TGTGAGGCCATTAGTGTGCTGTGAGGCCCTTGAGGAAGAGGCAAGACCTGTTTCAA 3300  
Db 3241 TGTGAGGCCATTAGTGTGCTGTGAGGCCCTTGAGGAAGAGGCAAGACCTGTTTCAA 3300  
QY 3301 GAGGCCATATGATATAGATGACATGTTACAAAAGGCTTCGGCTTGGTACACGTTTACATA 3360  
Db 3301 GAGGCCATATGATATAGATGACATGTTACAAAAGGCTTCGGCTTGGTACACGTTTACATA 3360  
QY 3361 TCATCCTACATATTTGGGTTGCTACATCAGGGGTTGAAAAGAGCTCATTTTCATTAGCTT 3420  
Db 3361 TCATCCTACATATTTGGGTTGCTACATCAGGGGTTGAAAAGAGCTCATTTTCATTAGCTT 3420  
QY 3421 TCCTGCTGCTGTTATGACACAGCTAATCCAGATTAAAGAGGACAAAGCACGTAACAGGCC 3480  
Db 3421 TCCTGCTGCTGTTATGACACAGCTAATCCAGATTAAAGAGGACAAAGCACGTAACAGGCC 3480  
QY 3481 AGTTCTCAACTGTCTATCTCAGGGCTCACTGAGTACAGATTAGTGTGTTGAAATGAGA 3540  
Db 3481 AGTTCTCAACTGTCTATCTCAGGGCTCACTGAGTACAGATTAGTGTGTTGAAATGAGA 3540  
QY 3541 TTCCAGTCCAGCGTTAAGCTGATATATATAATGATTAATAGGTTGATCATAGAANAAC 3600  
Db 3541 TTCCAGTCCAGCGTTAAGCTGATATATATAATGATTAATAGGTTGATCATAGAANAAC 3600  
QY 3601 TGTATGCAATGTTGACTACCTTTTGTCTTTAAACTGCAATGCAAGCTGCAACATATATGC 3660  
Db 3601 TGTATGCAATGTTGACTACCTTTTGTCTTTAAACTGCAATGCAAGCTGCAACATATATGC 3660  
QY 3661 AGTACTCTAAGAAACAGATGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 3720  
Db 3661 AGTACTCTAAGAAACAGATGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 3720  
QY 3721 TTTCTTCTAAA 3731  
Db 3721 TTTCTTCTAAA 3731

## RESULT 2

US-09-294-093B-1147  
; Sequence 1147, Application US/09294093B  
; Patent No. US20010051335A1  
; GENERAL INFORMATION:  
; APPLICANT: Lalqudi, Raghunath, V.  
; APPLICANT: Ito, Laura, Y.  
; APPLICANT: Sherman, Bradley, K.  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL  
; FILE REFERENCE: PL-0009 US  
; CURRENT APPLICATION NUMBER: US/09/294,093B  
; CURRENT FILING DATE: 1999-04-16  
; PRIOR APPLICATION NUMBER: 60/082,567  
; PRIOR FILING DATE: April 21, 1998  
; NUMBER OF SEQ ID NOS: 6207  
; SOFTWARE: PERL Program  
; SEQ ID NO 1147  
; LENGTH: 282  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700343843H1







```

; PRIOR APPLICATION NUMBER: 60/202,111
; PRIOR FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2720
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-847-665-3

Query Match      1.0%; Score 37.8; DB 10; Length 2720;
Best Local Similarity 44.9%; Pred. No. 4.9;
Matches 144; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 100 CTACTGCTGGGTAGTTTTATTTTTCATCACTTCCAGGGGTATTCAGTTGGTTAGCA 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2144 CTTCTGAAGCTCTTTTTCATCTGTCAGTCTATTAATCTGTGGCTATGAATCATCAGAAATGC 2203

QY 160 TTGGAAGTCGAAACGTCACCTTGGAAATTTGGCTACATGGGGAAGACAATTCAGTCTTTTCGG 219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2204 TAAGTCAGATCAATATTTTGTGGAAAAAAAATCTTGGGAAACAACCCAGGGTTTCGC 2263

QY 220 ATTCCTTATCTTCCTCTCGGGAAGTGGTTTAAGTCACTTCTAGAGAAATATACAGGATA 279
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2264 TGTGTGTGTTTTCCTTTCTTATTTTGTTTACTAGTCCCTTTAGCTAGTGATTAAT 2323

QY 280 TGGAACTGTATCTGCATCTGAGGTAAACAGTCCAAAGGAGGACTCTAGAGCATTTGCCAA 339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2324 TTGTTGTGCTCTCTTCATTTTGGCAATAACAATGCAATGAGATAATTAATACTTGGATGCTTA 2383

QY 340 AGTTCAATTTCCGACACACATAGTGTGCACAAAATCATCACCTTGGCTAAATAACAGGCT 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2384 AGAGGCTGCATATAGATAAGANTTCAGGCAAAACTACATTTATTTGTTTAATAACAGCTT 2443

QY 400 GTAATTTGGCTCTCTCTTAFTT 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2444 GTTCATAGGCTCTCTGTAATTT 2464

RESULT 11
US-09-847-665-6
; Sequence 6, Application US/09847665
; Patent No. US20020102586A1
; GENERAL INFORMATION:
; APPLICANT: ROGNER, UTE
; APPLICANT: SPYROPOULOS, DEMETRI
; APPLICANT: ROUGEULLE, CLAIRE
; APPLICANT: AVNER, PHILIP R
; TITLE OF INVENTION: IDENTIFICATION OF NEURAL DEFECTS ASSOCIATED WITH THE
; TITLE OF INVENTION: NUCLEOSOMAL ASSEMBLY PROTEIN 112 GENE
; FILE REFERENCE: 03495-0203-00000
; CURRENT APPLICATION NUMBER: US/09/847,665
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/202,111
; PRIOR FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 3699
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-847-665-6

```

	Query Match	1.0%	Score 37.8	DB 10	Length 3699
	Best Local Similarity	44.9%	Pred. No. 5.9		
	Matches 144	Conservative 0	Mismatches 177	Indels 0	Gaps 0
QY	100	CFACTGCTGGGTAGTTTTATTTCGATAACTTCAGGGGGTATCCAGTGGTGTTAGCA	159		
Db	3140	CTTCTGAAGTCCTTTTGTGATGTAGCTATTAACTGTGGGTATGAAATGATCAGAATGC	3199		
QY	160	TTTGAAGTCGAACTCGACTTGGAAATTGGGTACATGGGAAGACAAATTCAGTGGTTTCGG	219		

```

Db 3200 TAAGTGAATCAATATTGTTTGGAAAAAATACTTGGGAAACACCCCAAGGGTTTTCGC 3259
QY 220 ATTCCCTATTCTCTCTCTCGGGAAGTGGTTAAAGTCATTCTTTAGAGAAATATACAGGATA 279
Db 3260 TGTGTGTTGTTTCTTTCTTTTCTATTGTTTGTCTTACCTTTAGCTAGTGGATTTAATT 3319
QY 280 TGGAACTGTATGTCATTTGGAGTTAAACAGTCCAAAGGAGGATCTAGACCATTTGCCAA 339
Db 3320 TTGTTGTGCGCGCTTCATTTTTCGAATACAAATGAGTAGAATTTAAAACCTTTGGATGCTTA 3379
QY 340 AGTTCAAATTTGCCGACACATAAGTGTGACAAAATCATCACCTTTGGCTAATAACAGGCT 399
Db 3380 AGAGCCCTGCATATAGATAAGAAATTTCAGGCCAAAACACTATTATTGTTAATAACAGCTT 3439
QY 400 GTATTTGGCTCTCTCTATT 420
Db 3440 GTTCATAGGCTCTTGTATT 3460

RESULT 12
US-09-974-300-1446
; Sequence 1446, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1446
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1446

```

Query Match	1.0%	Score	37.6;	DB	10;	Length	567;
Best Local Similarity	46.8%	Pred.	No. 2.1;				
Matches	118;	Conservative	0;	Mismatches	134;	Indels	0;
				Caps			

  

Qy	1341	ATTATTCTGAAGACATAGATAACTTTCTTCGTTCTTTTGTGTGATGAGAGTGGGAGA	1400
Db			
Db	1	ATGCTTATTATGAAGATTAGATGTTCTCTTTCTTCGTCGATGACGTGCAGTCAAGCT	60
Qy	1401	AACTGTATTCTACAGACTTATTACAAAGCAAGTACTGGAAGTGGTGCAGGACAAACA	1460
Db			
Db	61	ATCTCGCAACAAAAGATTATATCAAAAGACATACACAAATTTGGTCTCATTTCAAAA	120
Qy	1461	TCTATGAGAGATCTTATCACTCGCGGAAGSCTTTGTAAATGGTGATAAAAAATTG	1520
Db			
Db	121	TGGATGATTTTACAGGAAGATCGAATGAAGSGGTATATAAAGCGCTTGGTGAAGCCA	180
Qy	1521	AATTTCTTGCATTTTCATCGACCGAGTTGGGGGATATATTCAGTGTGATGTTTCATCAA	1580
Db			
Db	181	AATTACGATTTTACCOCGAGCATGTGCTTTCGTACAATATAGACACGAAGCTGACCTAT	240
Qy	1581	GACCTGGSCCTTA	1592
Db			
Db	241	ACACAGACCTGA	252

  

RESULT 13

US-09-864-761-412/c

; Sequence 412, Application US/09864761

; Patent NO. US20020048763A1

; GENERAL INFORMATION:

RESULT 13  
US-09-864-761-412/c  
; Sequence 412, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeonica-X-1  
CURRENT APPLICATION NUMBER: US 09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annotax Sequence Listing Engine vers. 1.1  
SEQ ID NO 412  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC007543.4  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3  
OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 2.2  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.8  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2  
US-09-864-761-412

Query Match 1.0%; Score 36.4; DB 10; Length 499;  
Best Local Similarity 53.5%; Pred. No. 4.1;  
Matches 76; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 541 TACAGAGGTTCAATTGGCATTGGATTGGAAGAAATTTTCTTTTCTTTTCTAG 600  
DB 241 TGCACAGGATTAGTCTTTATAGCCATCAAAAATGATAGTTTTTCCATTGTTG 182

QY 601 TGGTTCAGCTGACTATAAAGCTTTCAGCTTTCATATAGAAATATATGGCAGGTGTGCTCCA 660  
DB 181 TTTTCTAAATGCTATATACATTAACCTCCAGCAACACTTTGTGGAAGGTTTAATCTA 122  
QY 661 TCGTCCATATGCTCAAAATGCT 682  
DB 121 CTACCCAGAAAGCTAGAGCT 100  
RESULT 14  
US-09-939-980-194  
Sequence 194. Application US/09939980  
Patent No. US20020082234A1  
GENERAL INFORMATION:  
APPLICANT: Black, Michael  
Burnham, Martin  
Hodgson, John  
Knowles, David  
Lonetto, Michael  
Nicholas, Richard  
Pratt, Julie  
Reichard, Richard  
Rosenberg, Martin  
Ward, Judith  
TITLE OF INVENTION: No. US20020082234A1 Polypeptides and Their Uses  
NUMBER OF SEQUENCES: 534  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/939,980  
FILING DATE: 27-Aug-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/936,165  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimmi, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 194:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3191 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 194:  
US-09-939-980-194

Query Match 1.0%; Score 36.4; DB 10; Length 3191;  
Best Local Similarity 49.0%; Pred. No. 13;  
Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 260 TTAGAAGATATACAGGATATGGAAGTCTATGTGCTTGGAGGTTAAACAGTCCAAAGGA 319  
DB 2823 TTAGAAGAACATTTATTTATGAAGATGTTGGACACCGGAGGATTTGAAAAAATAT 2882  
QY 320 GGATCTAGAGCAATTGCCAAAGTTCATTTGCCGACACATAAGTGTGACAAATCATC 379

Db 2883 CGNCTAATCGTGGCAATATATGGTCTGTCAGATAAAGAAACAAAGGATTT 2942  
Qy 380 ACTTGGCTAATACAGGCTGTATTTGGCTCTTCTTCTTATTTGAAGGCTTGGGAATGAAA 439  
Db 2943 ACCTTTCCTAAGAAAGTCAGTATTTGAAAACTTGACTTTAGTGGATCAGNAAT 3002  
Qy 440 ACTGATATTGTC 451  
Db 3003 CCGTGGTGGC 3014

RESULT 15

US-09-764-847-1401/c  
; Sequence 1401, Application US/09764847  
; Patent No. US20020132767A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC009  
; CURRENT APPLICATION NUMBER: US/09764,847  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 2003  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 1401  
; LENGTH: 1196  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-847-1401

Query Match 1.08; Score 36.2; DB 10; Length 1196;  
Best Local Similarity 50.9%; Pred. No. 8;  
Matches 138; Conservative 0; Mismatches 128; Indels 5; Gaps 2;  
Qy 3109 AGCTTTTCACTACAAAGCTGAATATGACAAAGCTGGGTAAATTAATGGACTACTATG- 3167  
Db 315 ACCAATTTTCATGAAAATGGAACTGTGAATTTGCTCTTTAAATCACTAGTCATCTCTGT 256  
Qy 3168 --GCATAAAACAGAGGCTGAAATACTAGTGGGCAATATGAAGGCATCAAAAACTTT 3225  
Db 255 TCATTTAAAAATAAAATATATAGTTCCTTAAGTTGTATGCTAAAGGCATCAGAAGAATT 196  
Qy 3226 TCACCGCAGAAAGATGCTGAGGCCATAGTGTGCTGTGAGGCCCTTGAGGAGGAGGC 3285  
Db 195 CATCTTAAGCAATAGCATGTGACTATTTTGCCTTATGAGATGGTGTATTTGGGACTAGG 136  
Qy 3286 AAGAGCCTGGTTCAAGAGCGCTAATGATAT--AGATGACATGTTACCAAGGCTTCGGCT 3343  
Db 135 ATGGCCATATATCAATATGTACAGGATATCGAGATCCAAATGTTAGCCAGAGATCTCAAT 76  
Qy 3344 TGGTACCAGGTACATATCATCCTACATATT 3374  
Db 75 TAGTCCCAAGTTTGTCTCTCATATTCATTTT 45

Search completed: November 5, 2002, 23:26:24  
Job time : 283 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 19:15:22 ; Search time 4534 Seconds  
(without alignments)  
13327.164 Million cell updates/sec

Title: US-09-782-874-1  
Perfect score: 3731

Sequence: 1 GAAATATCTTTTACTTACTT.....AGTTTCATCTTCTTCTTAA 3731

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em\_estba:\*

2: em\_esthm:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: gb\_estfun:\*

16: em\_estom:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_pln:\*

21: em\_gss\_vrt:\*

22: em\_gss\_fun:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	720.4	19.3	2199	11 AY103827	AY103827 Zea mays
2	662	17.7	773	14 BQ507474	BQ507474 EST614899
3	661.8	17.7	723	12 BG594667	BG594667 EST493345
4	625.4	16.8	659	13 B1203162	B1203162 EST521222
5	605	16.2	618	10 BE435332	BE435332 EST406410
6	603.2	16.2	682	12 BG597875	BG597875 EST496553

```

7 579 15.5 597 10 AM649811
8 565.2 15.1 610 12 BE920043
9 545.4 14.6 570 10 BE433646
10 529 14.2 537 10 AW932918
11 525 14.1 525 10 AW933692
12 482.6 12.9 521 10 BQ111608
13 473.6 12.7 550 9 AJ487433
14 458.8 12.3 529 12 BG592782
15 421 11.3 421 9 AI896465
16 402 10.8 434 14 BQ111609
17 388.6 10.4 403 12 BF052042
18 367.8 9.9 788 12 BG644922
19 365.4 9.8 367 12 BF112386
20 349 9.4 393 12 BG889665
21 334.8 9.0 688 14 BQ867964
22 329.8 8.8 443 10 AM623969
23 314.4 8.4 820 17 BH739145
24 308.4 8.3 688 10 AW587463
25 304.8 8.2 571 14 BQ488250
26 300.8 8.1 924 12 BG415065
27 295.8 7.9 623 12 BG526333
28 292.4 7.8 732 14 BQ967336
29 290.4 7.8 711 14 BQ967193
30 290.2 7.8 746 14 BQ970691
31 290 7.8 772 14 BQ967046
32 288.2 7.7 685 14 BQ971533
33 287 7.7 321 10 BE340733
34 272.2 7.3 756 14 BQ966709
35 269.4 7.2 617 14 BQ966840
36 265.2 7.1 629 13 BQ268864
37 264.8 7.1 588 14 BQ916374
38 264.4 7.1 741 14 BQ997497
39 262 7.0 473 13 B1130789
40 259.6 7.0 340 14 BQ509665
41 259 6.9 573 14 BQ786445
42 257.4 6.9 747 14 BQ970203
43 257 6.9 791 17 BH558894
44 253 6.8 793 14 BQ916044
45 249 6.7 786 17 BH467203

```

#### ALIGNMENTS

```

RESULT 1
AY103827
LOCUS Zea mays PC0135426 mRNA sequence.
DEFINITION Zea mays PC0135426 mRNA sequence.
ACCESSION AY103827
VERSION AY103827.1 GI:21206905
KEYWORDS HTC.
SOURCE Zea mays.
ORGANISM Zea mays.
REFERENCE 1 (bases 1 to 2199)
AUTHORS Hailey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitesitt,M.S.,
Arthur,L.W., Hanarey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 2199)
AUTHORS Coe,E.C.
TITLE Direct Submission
JOURNAL Submitted (25-Apr-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
FEATURES
source
1..2199
/organism="Zea mays"
/db_xref="MaizeDB:637927"
/db_xref="taxon:4577"
/clone="PC0135426"

```

2199 bp mRNA linear HTC 25-MAY-2002

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.



```

/clone.lib="Maize Mapping Project/DuPont Cornsensus
Library"
/notes="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"
550 a 435 c 530 g 584 t
BASE COUNT
ORIGIN

```

Query Match	19.3%	Score 720.4	DB 11	Length 2199
Best Local Similarity	64.1%	Pred. No. 2.3e-173		
Matches 1203	Conservative 0	Mismatches 501	Indels 72	Gaps 5
QY 1601	GATATAAGAGCTTGGATGGGTGATTTTTCGCAGATCAAGAAATGCGCAAAATATGCTGC	1660		
DB 3	GACATAAGAACTGGATGGGGACTTTCAGATATACGAATATGGCAAGATATGCTGCA	62		
QY 1661	AGACTTGGTCAATCTTTTGGTTCCTCCAGAGAGACTTTGAGTGTCTTAGCCATGAGATT	1720		
DB 63	ATACTTGGGCAATCTTTTCACTTCCACAGAAACTTTAAAGGTACACAATCTGAGGTG	122		
QY 1721	GAAGTTATTCCCGATGTAAAGGTTTCATGGCAACCAAGCTATGCTTTTGTGATGAAATG	1780		
DB 123	GAACGAATTCCTGATATTA--CAAAATGGCAAAAGTACATATCTCTGATGAGTTGGA	179		
QY 1781	AAAATATCTGGTGACTTTGCTCTCAPAGATTCGCTCAAAATGTGGCCCTCAA---TATAC	1837		
DB 180	AAATCTCAGCTAAATTTTGCAGTGGAGTGGCTATGAAGTGCAAAATGAAACGCTTGT	239		
QY 1838	CCATCTGCTTTCCAGATTCGTTATGGTGGATATAAAGGTGTGTGGTGTGATCCGGAT	1897		
DB 240	CCTTCTGTTTTTCAGATAAGGTATGGCGTTTACAAGGTGTGTGCTGTAGATACAAGA	299		
QY 1998	TCATCAATGAAGTGTCTTTGAGAAGAGCATGTGCAAAATATGATCAGACACATTAAG	1957		
DB 300	TCAATCATAAAGCTTTCTTTTGAGAAAGACATGTCAAAGTTCCAGTCCAGAAATATCA	359		
QY 1958	TTAGATGTCCTTGGATGGAGCAAAATATCAGCGCTTGTATCTTAACTGCTCAACTGAT	2017		
DB 360	CTTGATGTCCTGCATACAGCAAGTACCAACCATGCTTCTGAATCGGCAGTTGATTACT	419		
QY 2018	CTCTTGCTACACTTGGAGTGAAGATCAATCTTCGAAGCAGAGCAAGAGGAAGCTGTA	2077		
DB 420	CTTCTCTCAACACTGGGTTAGCCATATGCTTTGAGCTTAAGCAGAAGGAAGCCTTA	479		
QY 2078	GATCAGCTTGATGCTATCTTGATGATCTTTTGAAGGCACAGAGGCTTTGGAAATGATG	2137		
DB 480	AGGCAGTTGAACAGAATGTTAACTGAACACACAGGCTGCTGTGAAGCAGTTGAATATG	539		
QY 2138	TCCTCTGGAGACACTTAATATCTCAAGGCATGCTAAACGTGTGTATAGCCCTGAT	2197		
DB 540	CCATGGGAGAGTAACCAATATGATTAAGAATTTGTCATGTGCTACCAAGCCTGAT	599		
QY 2198	GCTGAGCCCTTCTTTTCAATGATGTGTGAACCTTCGCGCATCCAAAGTTGCTCGATTG	2257		
DB 600	CATGAGCCATATCTTTCCATGCTGTACAACTTTAGAGCATCCAAAGCTTCAGAGTTG	659		
QY 2258	CGGACTAGATCAAGATATTTATTCCAAATGGAGACAATGATGGATGTTTGGATGAA	2317		
DB 660	AAACAAAGTCAAGGATATTCATCACAGGGGGGAGCAATGATGGTGTGCCCTGGATGAA	719		
QY 2318	TCCAGAACCTTGGAAATATGGTCAGGTGTGTGTTCAGTTTACTGTGTCGACATGGAGAG	2377		
DB 720	ACCTGCACACTTAAGTACGCGCCAGGTATTCTGTCCAAGCTTCTTACAGTCGAGATG	774		
QY 2378	TTTTCTGACGATTTACATCCATTTAATACAGCAGATCCACACAGTAATTCATCTG	2437		
DB 775	-----ACCATCCAGGTCGGTGTGA794			
QY 2438	AAGGAAATGTGGTGTGTTCAAAAAATCCATGCTTGCATCCTGGTGATATTCGTGTTTA	2497		

Db	795	ACTGGAAGTAGTTGTCGCCAAAATCCTTGTCTCCACCCTGCTGCACATACAGGGTTCTC	854
Qy	2498	AAGCTGTAATATGTTTCGAGCGCTGCACACATAGGTAGATTTGTGTTATTTCCCTTCAGAAA	2557
Db	855	CAGCGTGTGATGTTTCTCTCTCTGCACCACCTTGTGTGACTGTGTGCTTTTCCACACGAG	914
Qy	2558	GGAAAAGACCTCATCCGAATCAATGTTCTGGAGTGATTTGGATGGGGATATCTACTTT	2617
Db	915	GGACCGCCGACCCCTAATGAGTGTTCAGGAGTGATCTTGATGGGACATATATTTT	974
Qy	2618	GTTTGTGGATCAAGACATGATCCCGCCCAAGGCAAGTCCAGCGGATGGATATCCTCCA	2677
Db	975	GTTTCTTGGATCCACATCTTATTCCAAAGTCGTTTGGTGGATCTCTAGGACTATCTCCA	1034
Qy	2678	GCACCCAGCATACAGTTGGACCATGATGTCACAAATTTAGGAAAGTTGAAGAGTACTTCACC	2737
Db	1035	GCITTCAGCAGAAACATTAGACCATGATGTCACATTTGAGGAGATACAGGAGTACTTCACA	1094
Qy	2738	AACATATTTGTAATCAGACAGTTTGGAAATCATGCAAAATGCCATGCTGPTATTTGGAGAC	2797
Db	1095	AACATCATAGTTAATGAGAGTCTTGGGATATCGGCAATGGCATGTGGTCTTTACAGAT	1154
Qy	2798	AGAGAACCTGATATGGCCATGAGTCATCCATGCAAAAAAAGTTGTCAGCTCTTTTCAATT	2857
Db	1155	CAGGAAGTATGAAAGCTGAGAGTCCACCGTCCGTTCAACTGGCCAGAGCTCTTCTCTATA	1214
Qy	2858	CGAGTGGACTTTCCAAAGACTGGTGTCCCGCTGAAATACCATTCTCAGTTGCGGCCATAA	2917
Db	1215	GCCTGTGATTTCCTCAAGACTGGAGTGGCGGCTCTGATTTCCACATGAGCTACATGTCAG	1274
Qy	2918	GAATACCCAGACTTCATGGATTAAGCGGACACAGACCAAGCTATATCTCAGAAAGAGTTATT	2977
Db	1275	GAGTATCCTGACTTCATGGAGAACTCGCAAAAGTCACTATGAATCAAGGGTGTGATC	1334
Qy	2978	GGAAAGCTTTTCAGGAAAGTGAAGGACAAGCAAGCTCAGGCTAGCTCTATGCGCAGCTTC	3037
Db	1335	GGGAAGCTCTATAGGAAATAAAGAGCAGCACACACAC-----ATAAGCACTCTT	1385
Qy	3038	ACAAGAGATGTTCAAGGAGATCATATGATGCTGATATGGAAGTTGATGATTTGAAGAT	3097
Db	1386	ACGAGGAACTGCCAAGCGGCTCTTATGACCCGATTTGATTGTGATGGCTATGAAGAT	1445
Qy	3098	TACATTCAGCAAGCTTTTGACTACAAAACCTGAATATGACACAAGCTGGGTAAATTTAATG	3157
Db	1446	TACATTACTGAGCTATAGATTCAAGGAAGATACGATTTTCAGGCTGGGTATCTTAATG	1505
Qy	3158	GACTACTATGGCATAAAACAGAGGCTGAAATACCTTAGTGGTCGATTAAGGCGATCA	3217
Db	1506	GACCACTATGGCATAAAGACTGAAGCTGAGATAAAGTGGATGATTTCTAAGATGGCA	1565
Qy	3218	AAAACCTTTGACCGCAGAAAAGATGCTGAGGCCATTTAGTGTCTGTGAGGGCCTTGAGG	3277
Db	1566	AAGAATTTTCAACCAAGATATGATGTGTATGCAATTAAGATGGCGGTGAGATCTTTGAGG	1625
Qy	3278	AAGGAGGCAAGAGCGCTGGTTTCAAGAGCGCTA-----ATGATATAGATGACATG	3325
Db	1626	AAAGAGCTAGTCTGGTGTTCATGAGATGACACAGGAGAGGATGGCCAAAGATGCCCAT	1685
Qy	3326	TTACCAAAGGCTTCGGCTTGGTACACAGTTACATATCATCTACATATTTGGGTTGCTAC	3385
Db	1686	GAGCCCAAGGCCCTCTGCTTGGTACCATGTTACTTATCATCAGCAGTACTGGGGCAGCTAC	1745
Qy	3386	AATCAGGGTTTGAAGAGAGCTCATTTCAATTAGCTTTCCTCTGGTGTATTATGACACGTA	3445
Db	1746	AATCAAGGGTATGATGGCGGCATCTATTAGCTCCCATGTTGCGTATATGACAAGCTT	1805
Qy	3446	ATCCAGATTTAAGAGG	3461
Db	1806	GTGGCCATCAAGCAGG	1821

RESULT 2  
BQ507474

```
LOCUS      B0507474              773 bp      mRNA      linear      EST 22-JUL-2002
DEFINITION EST614889 Generation of a set of potato cDNA clones for microarray
analyses mixed potato tissues Solanum tuberosum cDNA clone STMG059
5' end, mRNA sequence.
ACCESSION  B0507474
VERSION    B0507474.2  GI:21923304
KEYWORDS   EST.
SOURCE     potato.
ORGANISM   Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE  1 (bases 1 to 773)
            Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,
            Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and
            Karamycheva,S.A.
            Generation of a set of potato cDNA clones for microarray analyses
            Unpublished (2002)
            On Jun 10, 2002 this sequence version replaced gi:21366343.
            Contact: Robin Buell
            The Institute for Genomic Research
            9712 Medical Center Dr, Rockville, MD 20850, USA
            Email: potato@igr.org
            This clone is available through the Research Genetics, contact the
            Research Genetics for further information 1-800-711-6195 or
            cdna@resgen.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source           1..773
                     /organism="Solanum tuberosum"
                     /cultivar="Kennebec or Binjite"
                     /db_xref="taxon:4113"
                     /clone="STMG059"
                     /clone_lib="Generation of a set of potato cDNA clones for
                     microarray analyses mixed potato tissues"
                     /tissue_type="mixed tissues"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
                     XhoI; supplier: Combination of untreated and PhytoPhthora
                     infestans-treated libraries of stolons, leaves, leaflets,
                     axillary buds of stem explants, petioles, germinating eyes
                     tubers, or roots."
BASE COUNT  205 a 125 c 181 g 262 t
ORIGIN
Query Match      17.7%  Score 662;  DB 14;  Length 773;
Best Local Similarity 94.7%  Pred. No. 1.4e-158;
Matches 729;  Conservative 0;  Mismatches 35;  Indels 6;  Gaps 4;

QY  74  ATTGGTTGCTTCAGTTTCAGTCACACTACTGCTGGTAGTATTTATTTGCAATACTTC 133
DB  8  ATTGGTTGCTTCAGTTTCAGTTTCAGTCACACTACTGCTGGTAGTATTTATTTGCAATACTTC 67

QY  134 AGGGGATTCACAGTTGGTG-TTAGCATTTGAAAGTCGAAGTCACACTTGGAAATTTGGCTA 192
DB  68 AGGAGGTATTCAGTTGGTGATAGCATTTGAAAGTCGAAGTCACACTTGGAAATTTGGCTG 127

QY  193 CATGGAAAGACAATTCAGGTTTTCGGATTCCTCTATCTCTCTCTCGGAAGTGGTTAA 252
DB  128 CATGGAAAGACAATTCAGGTTTATGGATTCCTCTCTCTCTCTCGGAAGTGGTTAA 187

QY  253 GTCATTCTTAGAATAATACAGATATGGAAGTCGATGTCATGTCAGGTTAAACAGTC 312
DB  188 GTCATTCTTAGAATAATACAGATATGGAAGTCGATGTCATGTCAGGTTAAACAGTC 247

QY  313 CAAAGGAGGATCTAGAGCATTTGGCAAGTTCAATTTGGCCGACACAATAAGTGTGTGACAA 372
DB  248 CAAAGAGGATCTAGAGCATTTGGCAAGTTCAATTTGTGCAACAACATTAGTGTGTGACAA 307

QY  373 AATCATCACTTTGGCTAATACAGGCTGTATTTTGGCTCTCTCTTATTTGAAGGCTTGGGA 432
DB  308 AATCATCAATTTGGCTAGTAGAGGCTGTATTTTGGCTCTCTCTTATTTGAAGGCTTGGGA 367
```

```
QY  433 AATGAAACTGATATGTCCAACTGCGGGGCATATGTGGATCAGATGATGGCATACACTTT 492
DB  368 AAT--AACTGATCTTGTCCAACTGTGGGAGTATGTGGATCAGATGATGGCATACACTTT 424

QY  493 GAATTTCCGATGTCAGATATCAGATGACAAGTTTGCAGTGTGGGAAGTACAGAAGTTTC 552
DB  425 GAATTTCCGATGTCAGATATCAGATGACAAGTTTGCAGTGTGGGAAGTACAGAAGTTTC 484

QY  553 AATTCATTTGGCATTTGGATTTGAAGAAATTTTTTTTCTTTTATCTAGTGGTTCACTGA 612
DB  485 AATTAATTTGGCATTTGGATTTGAAGAAATTTTTTTTCTTTTATCTAGTGGTTCACTGA 544

QY  613 CTATTAACCTCAGCTTTTCATATCAATAATA-TGGCAGCTTGGTCCATCGTCCATATG 671
DB  545 TTATAACCTCAGCTTTTCATATCAATAATA-TATGGCAGGTTGGTCCATCGTCCATATG 604

QY  672 GTCAAAATGCTCAGTTTCTCCTCATACAGTATTTTGGTCTCTCTCGGATCTTAAGAGAC 731
DB  605 GTCAAAATGCTCAGTTTCTCCTCATACAGTATTTTGGTCTCTCTCGGATCTTAAGAGAC 664

QY  732 TTGAAAACCTCCTGTTTATAGCTTCTTTAAGAAACCTCTGATGATCAGTGGTGAGACAA 791
DB  665 TTGAAAACCTCCTGTTTATAGCTTCTTTAAGG-AACTCCTGATGATCAGTGGTGAGACAA 723

QY  792 CAGATTTCCCTCATCTTGGTAGGCTATCTTCTAGCTTATGTTTTCAG 841
DB  724 CAGATTTCCCTCATCTTGGTAGGCTCTCTCTAGCTTATGTTTTCAG 773

RESULT 3
BG594667      725 bp      mRNA      linear      EST 12-APR-2001
LOCUS      EST493345 cSTS Solanum tuberosum cDNA clone cSTS8119 5' sequence,
DEFINITION  mRNA sequence.
ACCESSION  BG594667
VERSION    BG594667.1  GI:13612807
KEYWORDS   EST.
SOURCE     potato.
ORGANISM   Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE  1 (bases 1 to 725)
            van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chiemiango,A.,
            Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
            Generations of ESTs from sprouting potato eyes
            Unpublished (2000)
            Contact: Cathy Ronning
            The Institute for Genomic Research
            For clone info: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdna@resgen.com
            Seq primer: M13F-R.
FEATURES             Location/Qualifiers
     source           1..725
                     /organism="Solanum tuberosum"
                     /cultivar="Kennebec"
                     /db_xref="taxon:4113"
                     /clone_lib="cSTS"
                     /tissue_type="sprouting eyes from tubers"
                     /dev_stage="12-14 weeks post harvest"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
                     XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
                     taken from tubers. The tubers were incubated at 26C in the
                     dark for 2-3 weeks prior to sprouting. The eyes were
                     frozen in liquid nitrogen immediately upon removal from
                     tubers."
BASE COUNT  228 a 140 c 170 g 187 t
ORIGIN
Query Match      17.7%  Score 661.8;  DB 12;  Length 725;
Best Local Similarity 96.4%  Pred. No. 1.5e-158;
```

Matches 699; Conservative 0; Mismatches 22; Indels 4; Gaps 2;

QY 2489 CGTGTTTAAAGGTGTAATGTTTCGAGCGCTGCACACATGTTAGATTTGGTGTATTC 2548  
|||||  
Db 2 CGAGTTTAAAGGCTGTAGATGTTCCAGCGCTGCACACATGTTAGATTTGGTGTATTC 61  
|||||

QY 2549 CCTCAGAAAGAAAAGACCTCATCCGAATGAATGTTCTGGAGTGAATTTGGATGGGAT 2608  
|||||  
Db 62 CCCAGAAAGAAAAGACCTCATCCGAATGAATGTTCTGGAGTGAATTTGGATGGGAT 121  
|||||

QY 2609 ATCTACTTTGTTCTGGATCAAGACATGATCCCGACAGGCAAGTCCAGCGATGAA 2668  
|||||  
Db 122 ATCTACTTTGTTCTGGATCAAGACATGATCCCGACAGGCAAGTCCAGCGATGAA 181  
|||||

QY 2669 TATCTCCAGACCCAGCATACAGTTGGACCATGATGTCACAAATGAGGAAGTTGAAGAG 2728  
|||||  
Db 182 TATCTCCAGACCCAGCATACAGTTGGATGATGTCACAAATGAGGAAGTTGAAGAG 241  
|||||

QY 2729 TACTTCACCACTATATGTAATGACAGTTTGGGAATCATAGCAAAATGCCATGCGTA 2788  
|||||  
Db 242 TACTTCACCACTATATGTAATGACAGTTTGGGAATCATAGCAAAATGCCATGCGTA 301  
|||||

QY 2789 TTTGCAGACAGAGAACCTGATATGCCATGATGATCCATGCAAAAAAATGCTGAGCTC 2848  
|||||  
Db 302 TTTGCAGACAGAGAACCTGATATGCCATGATGATCCATGCAAAAAAATGCTGAGCTC 361  
|||||

QY 2849 TTTTCAATGTCAGTGGACTTTCCAAAAGACTGGTGTTCCTCGCTGAAATACATCTCAGTTG 2908  
|||||  
Db 362 TTTTCAATGTCAGTGGACTTTCCAAAAGACTGGTGTTCCTCGCTGAAATACATCTCAGTTG 421  
|||||

QY 2909 CGCCTAAAGAAATACCCAGACTTCATGATAAGCGC--GACACAGACCATATATCTCA 2965  
|||||  
Db 422 CGCCTAAAGAAATACCCAGACTTCATGATAAGCGC--GACACAGACCATATATCTCA 481  
|||||

QY 2966 GAAAGAGTATTTGGAAGCTTTTCAGGAAAGTGAAGCAAAAGCACCTCAGGCTAGTCT 3025  
|||||  
Db 482 GAAAGAGTATTTGGAAGCTTTTCAGGAAAGTGAAGCAAAAGCACCTCAGGCTAGTCT 541  
|||||

QY 3026 ATCGCGACCTTCACAAGAGATTTGCAAGGAGATCATATGATGCTGATATGGAAGTTGAT 3085  
|||||  
Db 542 ACTGCGATCTTCACAGGAGATTTGCAAGGAGATCATATGATGCTGATATGGAAGTTGAT 601  
|||||

QY 3086 GGATTTGAAGATTACATTGACGAAGCTTTTGACTACAAAAGTGAATATGACAAACAGCTG 3145  
|||||  
Db 602 GGATTTGAAGATTACATTGACGAAGCTTTTGACTACAAAAGTGAATATGACAAACAGCTG 661  
|||||

QY 3146 GGTATTTAATGGACTACTATGGCATAAACACAGAGCTGAATACATAGTGGTGGCAT 3205  
|||||  
Db 662 GGTATTTAATGGACTACTATGGCATAAACACAGAGCTGAAT-CTAAGTGGTGGCATC 720  
|||||

QY 3206 ATGAA 3210  
|||||

Db 721 ATGAA 725  
|||||

RESULT 4  
BI203182  
LOCUS 659 bp mRNA linear EST 11-JUL-2001  
DEFINITION BI203182 cTOS Lycopersicon esculentum cDNA clone cTOS1C7 5' end similar to putative RNA-directed RNA polymerase, mRNA sequence.  
ACCESSION BI203182  
VERSION BI203182.1 GI:14680906  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.  
1 (bases 1 to 659)  
van der Hoeven,R., Sun,H., Bezzerides,J., Cho,J., Utterback,R., Ronning,C. and Tanksley,S.  
REFERENCE Generation of ESTs from Tomato Suspension Cultures

JOURNAL Unpublished (2001)  
COMMENT Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>.  
Location/Qualifiers  
1. .659  
/organism="Lycopersicon esculentum"  
/cultivar="TA496, E6203"  
/db\_xref="taxon:4081"  
/clone\_lib="cTOS1C7"  
/tissue\_type="suspension cultures"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; Suspension cultures of L.esculentum E6203 were grown in Murashige and Skoog based medium, supplemented with 15% coconut milk (filter sterilized and added after autoclaving), 2% saccharose, and lmg/ml 2,4D (pH5.8). Fresh medium was added every 7 days, and cultures were grown at 25 C, with 12hrs of light and continuous shaking."  
BASE COUNT 210 a 107 c 153 g 189 t  
ORIGIN

Query Match 16.8%; Score 625.4; DB 13; Length 659;  
Best Local Similarity 98.8%; Pred. No. 3.2e-149;  
Matches 651; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

QY 3060 CATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACATTTGACGAAGCTTTTGACT 3119  
|||||  
Db 1 CATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACATTTGACGAAGCTTTTGACT 60  
|||||

QY 3120 ACAAACCTGAATGACACACAGCTGGTAAATTTATGACTACTATGTCATAAAACAG 3179  
|||||  
Db 61 ACAAACCTGAATGACACACAGCTGGTAAATTTATGACTACTATGTCATAAAACAG 120  
|||||

QY 3180 AGCTGAATACATTAGTGGTGGCATTATGAAGSCATCAAAACTTTTGCCCGCAGAAAAG 3239  
|||||  
Db 121 AGCTGAATACATTAGTGGTGGCATTATGAAGSCATCAAAACTTTTGCCCGCAGAAAAG 180  
|||||

QY 3240 ATGCTGAGGCCATTAGTTGTTGCTGAGGCCCTTGAGGAGGAGGCAAGCGCTGTTCA 3299  
|||||  
Db 181 ATGCTGAGGCCATTAGTTGTTGCTGAGGCCCTTGAGGAGGAGGCAAGCGCTGTTCA 240  
|||||

QY 3300 AGAGGCTAATGATAGATGACATGTTACCAAGGCTTCGGCTTGCCTACCGCTTACAT 3359  
|||||  
Db 241 AGAGGCTAATGATAGATGACATGTTACCAAGGCTTCGGCTTGCCTACCGCTTACAT 300  
|||||

QY 3360 ATCATCTACATATTTGGGTTGCTACAATCAGGGGTTGAAAAGAGCTCATTTTCATTAGCT 3419  
|||||  
Db 301 ATCATCTACATATTTGGGTTGCTACAATCAGGGGTTGAAAAGAGCTCATTTTCATTAGCT 360  
|||||

QY 3420 TTCCTGGTGTGTTATGACCAGCTAATCCAGATTGAAGGACAAAGCAGCAAGGC 3479  
|||||  
Db 361 TTCCTGGTGTGTTATGACCAGCTAATCCAGATTGAAGGACAAAGCAGCAAGGC 420  
|||||

QY 3480 CAGTTCTCAACTTGTCTATCTCTCAGGCTCAACTGAGTCACAGATTAGTGTGAAATGAG 3539  
|||||  
Db 421 CAGTTCTCAACTTGTCTATCTCTCAGGCTCAACTGAGTCACAGATTAGTGTGAAATGAG 480  
|||||

QY 3540 ATTCCAGTCGAGCGTTAAGCTGATATATATATATGTAATGATAGGCTGATCATAGAAA 3599  
|||||  
Db 481 ATTCCAGTCGAGCGTTAAGCTGATATATATATATGTAATGATAGGCTGATCATAGAAA 540  
|||||

QY 3600 CTGTTATGCAATTTGTCAGTACCTTTTCTCTTTAAACTGTCATGAAGCTCAACATATATG 3659  
|||||  
Db 541 CTGTTATGCAATTTGTCAGTACCTTTTCTCTTTAAACTGTCATGAAGCTCAACATATATG 599  
|||||

QY 3660 CAGTACTCTTAAGAAACAGATGTACAGCTAGTACTAATATGATGTGATTGAGTTTCA 3718  
|||||  
Db 600 CAGTACTCTTAAGAAAC-GATGTACAGCTAGTACTAATATGATGTGATTGAGTTTCA 657  
|||||

RESULT_5	B2433932	618 bp	mRNA	linear	EST 18-MAY-2001
BEA35332	EST406410	tomato breaker fruit,	TIGR Lycopersicon esculentum	cDNA	
LOCUS	clone CLEG26C9,	mRNA sequence.			
DEFINITION	B2435332				
ACCESSION	B2435332.1	GI:9433175			
VERSION	EST.				
KEYWORDS	tomato.				
SOURCE	Lycopersicon esculentum				
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
	Asteridae; euasterids I; Solanales; Solanaceae; Solanum;				
	Lycopersicon.				
REFERENCE	1 (bases 1 to 618)				
AUTHORS	Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,				
	Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Renning,C.M.,				
	Nierman,W., Fraser,C.N., Martin,G.B., Giovannoni,J.J. and Tanksley				
	,S.D.				
TITLE	Generation of ESTs from tomato fruit tissue, breaker stage				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: CUGI				
	Clemson University Genomics Institute				
	Clemson University				
	100 Jordan Hall, Clemson, SC 29634, USA				
	Email: http://www.genome.clemson.edu/orders/index.html				
	5 prime sequence.				
FEATURES	Location/Qualifiers				
source	1..618				
	/organism="Lycopersicon esculentum"				
	/cultivar="TA496"				
	/db_xref="taxon:4081"				
	/clone="cLEG26C9"				
	/clone_lib="tomato breaker fruit, TIGR"				
	/tissue_type="pericarp"				
	/dev_stage="breaker"				
	/lab_host="SOLR"				
	/note=Vector: pBluescriptSKmCUadapt; Site_1: EcoRI;				
	Site_2: XhoI; Fruit were harvested at the breaker stage				
	(first sign of lycopene accumulation on the blossom end of				
	the fruit). Fruit were cut in half and the seeds and				
	locules were discarded prior to freezing the pericarp."				
BASE COUNT	196 a 115 c 149 g 157 t				
ORIGIN					
	Query Match 16.2%; Score 605; DB 10; Length 618;				
	Best Local Similarity 99.7%; Pred.No. 5.4e-144;				
	Matches 616; Conservative 0; Mismatches 1; Indels 1; Gaps 1;				
QY	2801	GACCTGATATGCCATGAGTGATCCCATG-CAAAAAACTTGCCTGAGCTCTTTCAATTGC	2859		
Db	1				
	1	GACCTGATATGCCATGAGTGATCCCATGCGAAAAACTTGCCTGAGCTCTTTCAATTGC	60		
QY	2860	AGTGGACTTTCCAAGAAGCTGGTGTTCGCCGTGAATACCATCTCAGTTCGCCCTCAAGA	2919		
Db	61	AGTGGACTTTCCAAGAAGCTGGTGTTCGCCGTGAATACCATCTCAGTTCGCCCTCAAGA	120		
QY	2920	ATACCCAGACTTCATGATAAGCGGCAGACAGACAGCTATATCTCAGAAAGATTATGG	2979		
Db	121	ATACCCAGACTTCATGATAAGCGGCAGACAGACAGCTATATCTCAGAAAGATTATGG	180		
QY	2980	AAAAGCTTTTCAGGAAAGTAGGACAAAGCACCTCAGGCTAGCTCTATCGCGACCTTCAC	3039		
Db	181	AAAAGCTTTTCAGGAAAGTAGGACAAAGCACCTCAGGCTAGCTCTATCGCGACCTTCAC	240		
QY	3040	AAGAGATGTTGCAAGGAGATCATATGCTGATATGGAAGTTGATGGATTGAAGATTGA	3099		
Db	241	AAGAGATGTTGCAAGGAGATCATATGCTGATATGGAAGTTGATGGATTGAAGATTGA	300		
QY	3100	CATTGACGAAAGCTTTTGACTTACAAAACCTGAATATGACAACAGCTGGTAAATTAATGGA	3159		



REFERENCE 1 (bases 1 to 610)  
 AUTHORS van der Hoeven,R.S., Bezzerides,J., Holt,I.E., Liang,F., Cho,J.,  
 Otterback,R., Hansen,C.L., Doan,B., Bougri,O., Buell,C.R., Ronning  
 C.M., Fry,W.E., Tanksley,S.D. and Baker,B.  
 TITLE Generation of ESTs from potato leaves and petioles  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Cathy Ronning  
 The Institute for Genomic Research  
 For clone request: please contact Research Genetics, Libraries  
 Division tel 1-800-711-6195, email cdnaresgen.com.  
 Location/Qualifiers  
 1..610

FEATURES  
 source

/organism="Solanum tuberosum"  
 /cultivar="Kennebec"  
 /db\_xref="taxon:4113"  
 /clone="CSTB3D19"  
 /clone\_lib="potato leaves and petioles"  
 /tissue\_type="leaflets and petioles"  
 /dev\_stage="8 weeks old plants"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:  
 XhoI; Tissue was supplied by Dr. Fry (Cornell University).  
 Leaflets and petioles were isolated from 8 week old  
 greenhouse grown plants. The plants were watered and  
 fertilized freely. The tissue was immediately frozen in  
 liquid nitrogen."  
 181 a 121 c 146 g 162 t

Query Match 15.1%; Score 565.2; DB 12; Length 610;  
 Best Local Similarity 96.6%; Pred. No. 8.7e-134;  
 Matches 589; Conservative 0; Mismatches 18; Indels 3; Gaps 1;

QY 2480 GGTGATATTCGTTTAAAGCTGTAATGTCGAGCGCTGCACCATGCTGATGTTGT 2539  
 DB 1 GGTGATATTCGTTTAAAGCTGTAATGTCGAGCGCTGCACCATGCTGATGTTGT 60  
 QY 2540 GTTGATATTCCTCAGAAAGGAAAGAGCTATCCGATGATGTTCTGGGAGTATTG 2599  
 DB 61 GTTGATATTCCTCAGAAAGGAAAGAGCTATCCGATGATGTTCTGGGAGTATTG 120  
 QY 2600 GATGGGATATCTACTTTGTTGGGATCAAGACATGATCCGCCCAAGCAATCCAG 2659  
 DB 121 GATGGGATATCTACTTTGTTGGGATCAAGACATGATCCGCCCAAGCAATCCAG 180  
 QY 2660 CCGATGGAATATCTCCAGACACCCAGCATACAGTTGGACCATGATGTCACAAATGAGAA 2719  
 DB 181 GCGATGGAATATCTCCAGACACCCAGCATACAGTTGGACCATGATGTCACAAATGAGAA 240  
 QY 2720 GTTGAAGACTACTTCACCAACTATATTGTGATGACAGTTTGGGATCATAGCAATGCC 2779  
 DB 241 GTTGAAGACTACTTCACCAACTATATTGTGATGACAGTTTGGGATCATAGCAATGCC 300  
 QY 2780 CATGTCGATTTGCGACAGAGAACTGATATGCGCATGAGTGATCCATGCAAAAACCTT 2839  
 DB 301 CATGTCGATTTGCGACAGAGAACTGATATGCGCATGAGTGATCCATGCAAAAACCTT 360  
 QY 2840 GCTGAGCTCTTTCATGCTGACGTGACATTCGAAAGCTGGTTCGCCCTGAATACCA 2899  
 DB 361 GCTGAGCTCTTTCATGCTGACGTGACATTCGAAAGCTGGTTCGCCCTGAATACCA 420  
 QY 2900 TCTCAGTTCCGCCCTTAAAGATPACCCAGACTTCATGGGATAGCCGCG--GACAAAGACCAAGC 2956  
 DB 421 TCTCAGTTCCGCCCTTAAAGATPACCCAGACTTCATGGGATAGCCGCGCGCAAGACACAC 480  
 QY 2957 TATATCTCAGAAGAGTTATTTGAAAGCTTTTCAGGAAGTGAAGGACAAAGACCTCAG 3016  
 DB 481 TATATCTCAGAAGAGTTATTTGAAAGCTTTTCAGGAAGTGAAGGACAAAGACCTCAG 540  
 QY 3017 GCTAGCTCTATCGACCTTCACAGAGATGTTGCAAGGAGATCATATGATGCTGATATG 3076  
 DB 541 GCTAGCTCTATCGACCTTCACAGGAGATGTTGCAAGGAGATCATATGATGCTGATATG 600

QY 3077 GAAGTTGATG 3086  
 DB 601 GAAGTTGATG 610

RESULT 9  
 BE433646 570 bp mRNA linear EST 18-MAY-2001  
 LOCUS EST400175 tomato breaker fruit, TIGR Lycopersicon esculentum CDNA  
 DEFINITION clone CUEGL6J9, mRNA sequence.  
 ACCESSION BE433646  
 VERSION BE433646.1 GI:9431489  
 KEYWORDS EST.  
 SOURCE tomato.  
 ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
 Lycopersicon.

REFERENCE 1 (bases 1 to 570)  
 AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,  
 Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,  
 Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley  
 S.D.  
 TITLE Generation of ESTs from tomato fruit tissue, breaker stage  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: CUGI  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Email: <http://www.genome.clemson.edu/orders/index.html>  
 5 prime sequence.

FEATURES  
 source  
 1..570  
 /organism="Lycopersicon esculentum"  
 /cultivar="RA498"  
 /db\_xref="taxon:4081"  
 /clone="CUEGL6J9"  
 /tissue\_type="tomato breaker fruit, TIGR"  
 /tissue\_type="Pericarp"  
 /dev\_stage="breaker"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescript-SKmuadapt; Site 1: EcoRI;  
 Site 2: XhoI; Fruit were harvested at the breaker stage  
 (first sign of lycopene accumulation on the blossom end of  
 the fruit). Fruit were cut in half and the seeds and  
 locules were discarded prior to freezing the pericarp."  
 157 a 112 c 134 g 167 t

Location/Qualifiers

Query Match 14.6%; Score 545.4; DB 10; Length 570;  
 Best Local Similarity 99.5%; Pred. No. 1e-128; 1; Indels 2; Gaps 2;  
 Matches 568; Conservative 0; Mismatches 2; Gaps 2;

QY 2137 GTCTCCTGGAGAGAACACTAATATTCGAAGGCAATGCTAAACTGTGGTTATAGCCTGA 2196  
 DB 1 GTCTCCTGGAGTGAACACTAATATTCGAAGGCAATGCTAAACTGTGGTTATAGCCTGA 60  
 QY 2197 TCGTGAGCCCTTCTTTCATGATGTTGCAAACTCCGGG-CATCCAAAGTGTCTCGATT 2255  
 DB 61 TCGTGAGCCCTTCTTTCATGATGTTGCAAACTCCGGGCAATGCTAAAGTGTCTCGATT 120  
 QY 2256 TCGGACTAGATCAAGAATATTTATCCAAATGGAAGCAATGATGGGATGTTGGATG 2315  
 DB 121 TCGGACTAGATCAAGAATATTTATCCAAATGGAAGCAATGATGGGATG-TTGGATG 179  
 QY 2316 AATCCGAACCTTGGAAATATGGTCAGGTGTTGTTCAGTTTACTGGTGGACATGGAG 2375  
 DB 180 AATCCGAACCTTGGAAATATGGTCAGGTGTTGTTCAGTTTACTGGTGGACATGGAG 239  
 QY 2376 AGTTTCTGAGGATTTACATCCATTTAATACAGCAGATCCCAACAGTAATTTCTATC 2435

```

Db 240 AGTTTCTGACGATTACATCCATTTAATAACACGAGATCCACACAGTAATTCATTC 299
QY 2436 TGAAGGGAATGTGGTTGTTCCAAAAATCCATCTTGCAATCCCTGGTGATATCGTGTTT 2495
Db 300 TGAAGGGAATGTGGTTGTTCCAAAAATCCATCTTGCAATCCCTGGTGATATCGTGTTT 359
QY 2496 TAAAGGCTGTAATGTTCGAGCGCTGCACACATGAGTAGATGTTGGTTGTTATCCCTCAGA 2555
Db 360 TAAAGGCTGTAATGTTCGAGCGCTGCACACATGAGTAGATGTTGGTTGTTATCCCTCAGA 419
QY 2556 AAGGAAAAGACCTCATCCGAATGAATGTTCTGGAGTGATTTGGATGGGATATCTACT 2615
Db 420 AAGGAAAAGACCTCATCCGAATGAATGTTCTGGAGTGATTTGGATGGGATATCTACT 479
QY 2616 TTGTTTGTGGATCAAGACATGATCCGCAAGCAAGTCCAGCCGATGGAATATCCTC 2675
Db 480 TTGTTTGTGGATCAAGACATGATCCGCAAGCAAGTCCAGCCGATGGAATATCCTC 539
QY 2676 CAGCACCAGCATACAGTTGGACCATGATGT 2706
Db 540 CAGCACCAGCATACAGTTGGACCATGATGT 570

RESULT 10
AW932918
LOCUS EST358761 tomato fruit mature green, TAMU Lycopersicon esculentum
DEFINITION cDNA clone cLEF50H7 5', mRNA sequence.
ACCESSION AW932918
VERSION AW932918.1 GI:8108319
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 537)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M.,
Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue
Unpublished (1999)
Contact: CUGI
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
Location/Qualifiers
1..537
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEF50H7"
/clone_lib="tomato fruit mature green, TAMU"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLEF - fruit were tagged at the 1cm stage and
harvested 3-5 days prior to ripening. Fruit were cut in
half to verify the seeds were indeed 'immature' and the
seeds and locules were discarded prior to freezing the
pericarp"
BASE COUNT 177 a 112 c 118 g 130 t
ORIGIN

Query Match 14.2%; Score 529; DB 10; Length 537;
Best Local Similarity 99.1%; Pred. No. 1.6e-124;
Matches 532; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2635 CATGATCCGCGCAAGGCAAGTCCAGCGATGGAATATCTCCAGCACCCAGCATACAGTT 2694

```

```

Db 1 CATGATCCGCGCAAGGCAAGTCCAGCGATGGAATATCTCCAGCACCCAGCATACAGTT 60
QY 2695 GGACCATGATGTGACAAATGAGGAAGTTGAAGAGTACTTCCACCACTATATTGTGAATGA 2754
Db 61 GGACCATGATGTGACAAATGAGGAAGTTGAAGAGTACTTCCACCACTATATTGTGAATGA 120
QY 2755 CAGTTTGGGAATCATAGCAAAATGCCCATGTCGTATTGTCAGACAGAGAACCCTGATATGCG 2814
Db 121 CAGTTTGGGAATCATAGCAAAATGCCCATGTCGTATTGTCAGACAGAGAACCCTGATATGCG 180
QY 2815 CATGAGTGATCCATGCAAAAAACTTCTGCTGAGTCTTTTCAATTGCAAGTGAATTTCCAAA 2874
Db 181 CATGAGTGATCCATGCAAAAAACTTCTGCTGAGTCTTTTCAATTGCAAGTGAATTTCCAAA 240
QY 2875 GACTGTGTTCGCGTGAATACCATCTCAGTTGCGCCCTTAAGAAATACCCAGACTTCAT 2934
Db 241 GACTGTGTTCGCGTGAATACCATCTCAGTTGCGCCCTTAAGAAATACCCAGACTTCAT 300
QY 2935 GGATAAGCCGGACAAGCAGCATATATCTCAGAAAGAGTTATGGAAAGCTTTTCAGGAA 2994
Db 301 GGATAAGCCGGACAAGCAGCATATATCTCAGAAAGAGTTATGGAAAGCTTTTCAGGAA 360
QY 2995 AGTSAAGGACAAAGCACCTCAGGCTAGCTCTATCGCGACCTTCACAAGAGATGTTGCAAG 3054
Db 361 AGTSAAGGACAAAGCACCTCAGGCTAGCTCTATCGCGACCTTCACAAGAGATGTTGCAAG 420
QY 3055 GAGATCATATGATGCTGATATGGAAGTTGATGGAATTTGAAGATTACATTGACGAAGCTTT 3114
Db 421 GAGAACATATGATGCTGATATGGAAGTTGATGGAATTTGAAGATTACATTGACGAAGCTTT 480
QY 3115 TGACTACAAACTGAATATGACAAACAGCTGGGTAAATTTAATGAGTACTATATGGCAT 3171
Db 481 TGACTACAAACTGAATATGACAAACAGCTGGGTAAATTTAATGAGTACTATATGGCAT 537

RESULT 11
AW933692
LOCUS EST359535 tomato fruit mature green, TAMU Lycopersicon esculentum
DEFINITION cDNA clone cLEF56C18 5', mRNA sequence.
ACCESSION AW933692
VERSION AW933692.1 GI:8109093
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 525)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M.,
Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
Location/Qualifiers
1..525
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEF56C18"
/clone_lib="tomato fruit mature green, TAMU"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:

```

Xhol; cLEF - Fruit were tagged at the 1cm stage and harvested 3-5 days prior to ripening. Fruit were cut in half to verify the seeds were indeed 'immature' and the seeds and locules were discarded prior to freezing the pericarp"

```
BASE COUNT 151 a 111 c 121 g 142 t
ORIGIN

Query Match 14.1%; Score 525; DB 10; Length 525;
Best Local Similarity 100.0%; Pred. No. 1.7e-123;
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2464 TCCATGCTTGCATCCTGGTGATATCGTGTGTTAAAGCGCTGAATATTCGAGCGTGCA 2523
Db 1 TCCATGCTTGCATCCTGGTGATATCGTGTGTTAAAGCGCTGAATATTCGAGCGTGCA 60
QY 2524 CCACATGCTAGATGCTGTGTTATCCCTCAGAAAGAAAAGACCTCATCCGAATGAATG 2583
Db 61 CCACATGCTAGATGCTGTGTTATCCCTCAGAAAGAAAAGACCTCATCCGAATGAATG 120
QY 2584 TTCTGGGAGTGATTTGGATGGGATATCTACTTTGTTGCTGGATCAAGACATGATCCC 2643
Db 121 TTCTGGGAGTGATTTGGATGGGATATCTACTTTGTTGCTGGATCAAGACATGATCCC 180
QY 2644 GCCAAGGCAAGTCCAGCGGATGGAATATCTCCAGCAGCAGCAGCAGCAGTGGACCATGA 2703
Db 181 GCCAAGGCAAGTCCAGCGGATGGAATATCTCCAGCAGCAGCAGCAGTGGACCATGA 240
QY 2704 TGTCAAAATTCAGGAAGTTGAAGAGTACTTCCACCACTATATTTGTAATGACAGTTGGG 2753
Db 241 TGTCAAAATTCAGGAAGTTGAAGAGTACTTCCACCACTATATTTGTAATGACAGTTGGG 300
QY 2754 AATCATAGCAATGCCATGCGTATTTGCAGACAGAGAACCTGATATGGCCATGAGTGA 2823
Db 301 AATCATAGCAATGCCATGCGTATTTGCAGACAGAGAACCTGATATGGCCATGAGTGA 360
QY 2824 TCCATGCAAAAACCTGCTGAGCTCTTTTCAATTCAGTGGACATTTCCAAAGAGCTGTGT 2883
Db 361 TCCATGCAAAAACCTGCTGAGCTCTTTTCAATTCAGTGGACATTTCCAAAGAGCTGTGT 420
QY 2884 TCCGCTGAATACCATCTCAGTTGCGCCCTAAAGATACCCAGCTTCATGGATAAGCC 2943
Db 421 TCCGCTGAATACCATCTCAGTTGCGCCCTAAAGATACCCAGCTTCATGGATAAGCC 480
QY 2944 GGACAAGACCAAGCTATATCTCAGAAAGAGTTATTGAAAGCTTTT 2988
Db 481 GGACAAGACCAAGCTATATCTCAGAAAGAGTTATTGAAAGCTTTT 525
```

```
RESULT 12
BQ111608
LOCUS
DEFINITION BQ111608 521 bp mRNA linear EST 22-JUL-2002
5' end, mRNA sequence.
Solanum tuberosum cdna clone STMCB87
```

```
ACCESSION BQ111608
VERSION BQ111608.2 GI:21915035
KEYWORDS EST.
```

```
SOURCE Solanum tuberosum
```

```
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
```

```
REFERENCE 1 (bases 1 to 521)
AUTHORS Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,
Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and
Karamycheva,S.A.
```

```
TITLE Generation of a set of potato cDNA clones for microarray analyses
JOURNAL Unpublished (2002)
```

```
COMMENT On Apr 17, 2002 this sequence version replaced gi:20163570..
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato@tigr.org
```

This clone is available through the Research Genetics, contact the Research Genetics for further information 1-800-711-6195 or cdna@resgen.com

Seq primer: T3.

#### FEATURES

##### Source

```
1..521
/organism="Solanum tuberosum"
/cultivar="Kennebec or Binjete"
/db_xref="taxon:4113"
/clone="STMCB87"
/tissue_lib="mixed potato tissues"
/tissue_type="mixed tissues"
/lab_host="SOL8"
/notes="Vector: pBluescript SK(-); Site:1: EcoRI; Site:2:
XhoI; supplier: Combination of untreated and Phytophthora
infestans-treated libraries of stolons, leaves, leaflets,
axillary buds of stem explants, petioles, germinating eyes
tubers, or roots."
BASE COUNT 155 a 106 c 121 g 139 t
ORIGIN
```

```
Query Match 12.9%; Score 482.6; DB 14; Length 521;
Best Local Similarity 96.7%; Pred. No. 1.3e-112;
Matches 504; Conservative 0; Mismatches 14; Indels 3; Gaps 1;
```

```
QY 2480 GGTGATATTCGTGTTTAAAGCGCTGTAATGTTTCGAGCGCTGCACACATGTTAGATTGT 2539
```

```
Db 1 GGTGATATTCGTGTTTAAAGCGCTGTTAGATGTTCCAGCGCTGCACACATGTTAGATTGT 60
```

```
QY 2540 GTTGTATTCCTCAGAAAGAAAAGACCTCATCCGAATGAATGTTCTGGGAGTGATTGG 2599
```

```
Db 61 GTTGTATTCCTCAGAAAGAAAAGACCTCATCCGAATGAATGTTCTGGGAGTGATTGG 120
```

```
QY 2600 GATGGGATATCTACTTTGTTGCTGGATCAAGACATGATCCGCAAGCAAGTCCAG 2659
```

```
Db 121 GATGGGATATCTACTTTGTTGCTGGATCAAGACATGATCCGCAAGCAAGTCCAG 180
```

```
QY 2660 CCGATGGAATATCCTCCAGCAGCAGCAGTGGACCATGATGTCACATTTAGGAA 2719
```

```
Db 181 CCGATGGAATATCCTCCAGCAGCAGCAGTGGACCATGATGTCACATTTAGGAA 240
```

```
QY 2720 GTTGAAGAGTACTTCCAACTATATTGTGATGACAGTTTGGGAATCATAGCAATGCC 2779
```

```
Db 241 GTTGAAGAGTACTTCCAACTATATTGTGATGACAGTTTGGGAATCATAGCAATGCC 300
```

```
QY 2780 CATGCTGATTTGCAGACAGAGAACCTGATATGGCCATGAGTGATCCATGCAAAAACCTT 2839
```

```
Db 301 CATGCTGATTTGCAGACAGAGAACCTGATATGGCCATGAGTGATCCATGCAAAAACCTT 360
```

```
QY 2840 GCTGAGCTCTTTCAATTCGAGTGGACTTCCAAAGACATGGTTCCTCGTGAATACCA 2899
```

```
Db 361 GCTGAGCTCTTTCAATTCGAGTGGACTTCCAAAGACATGGTTCCTCGTGAATACCA 420
```

```
QY 2900 TCTCAGTTGGCCCTAAAGAAATACCCAGACTTTCATGGATAAGCCG---GACAAGACGAC 2956
```

```
Db 421 TCTCAGTTGGCCCTAAAGAAATACCCAGACTTTCATGGATAAGCCG---GACAAGACGAC 480
```

```
QY 2957 TATATCTCAGAAAGAGTTATTGAAAGCTTTTCAGAAAGT 2997
```

```
Db 481 TATATCTCAGAAAGAGTTATTGAAAGCTTTTCAGAAAGT 521
```

#### RESULT 13

AJ487433

LOCUS

DEFINITION

AJ487433

ACCESION

VERSION

KEYWORDS

SOURCE

ORGANISM

potato.

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

AJ487433 550 bp mRNA linear EST 30-JUL-2002  
AJ487433 Solanum tuberosum cv. Provita Solanum tuberosum cdna clone  
P3f8, mRNA sequence.

AJ487433

VERSION AJ487433.1 GI:22022186

EST

potato.

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;



REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 550)	Comparative mapping between potato ( <i>Solanum tuberosum</i> ) and <i>Arabidopsis thaliana</i> reveals structurally conserved domains and ancient duplications in the potato genome	Unpublished (2002)	
2	Gebhardt, C., Walkemeier, B., Henselewski, H., Barakat, A., Delseny, M. and Stuber, K.	Plant breeding and yield physiology	Contact: Gebhardt C	
3	Carl-von-Linne Weg 10, Cologne, D-50829, Germany.	Location/Qualifiers		
4	1. .550	/organism="Solanum tuberosum"		
5	/cultivar="Provita"	/db_xref="taxon:4113"		
6	/map="v"	/clone="P3f8"		
7	/clone_lib="Solanum tuberosum cv. Provita"			
8	BASE COUNT 182 a 91 c 125 g 152 t			
9	ORIGIN			
10	Query Match 12.7%; Score 473.6; DB 9; Length 550;			
11	Best Local Similarity 93.1%; Pred. No. 2.7e-110;			
12	Matches 512; Conservative 0; Mismatches 24; Indels 14; Gaps 1;			
13	QY 3120	ACAAACTGATATGACAAACAGCTGGGTAAATTAATGGACTACTATGGCATAAAACAG	3179	
14	Db 1	ACAAACTGATATGACAAACAGCTGGGTAAATTAATGGACTACTATGGCATAAAACGG	60	
15	QY 3180	AGGCTGAATACTTAGTGGTGGCATTTGAAGGCATCAAAAACCTTTGACCGCAGAAAAG	3239	
16	Db 61	AGGCTGAATACTTAGTGGTGGCATTTGAAGGCATCAAAAACCTTTGACCGCAGAAAAG	120	
17	QY 3240	ATGCTGAGGCCATTAGTGTTCCTGTGAGGGCCCTTGAGGAAGGAGCGCAAGACCTGGTTCA	3299	
18	Db 121	ATGCTGAGGCCATTAGTGTTCCTGTGAGGGCCCTTGAGGAAGGAGCGCAAGACCTGGTTCA	180	
19	QY 3300	AGAGCGCTAATGATATAGATGACATGTTACCAAGGCTTCGGCTTGGTACCACGTTACAT	3359	
20	Db 181	AGAGCGCTAATGATATAGATGACATGTTACCAAGGCTTCGGCTTGGTACCACGTTACAT	240	
21	QY 3360	ATCATCTTACATATTTGGGGTTGCTACAATCAGGGGTTGAAAAGAGCTCATTTTCATTAGCT	3419	
22	Db 241	ATCATCTTACATATTTGGGGTTGCTACAATCAGGGGTTGAAAAGAGAGATCATTTTCATTAGCT	300	
23	QY 3420	TTCCCTGTGTGTTATGACCAAGCTAATCCAGATTAACAGGCAACAGCAGTAAACAGGC	3479	
24	Db 301	TTCCCTGTGTGTTATGACCAAGCTAATCCAGATTAACAGGCAACAGCAGTAAACAGGC	360	
25	QY 3480	CAGTTCTCAACTTGTCATCTCTCAGGGGCTCAACTGAGTCACAGATTTAGTGTGAAATGAG	3539	
26	Db 361	CAGTTCTCAANTTGTCACTCTCAGGGGCTCAGTGAGTCGCAAAATAGTGTGAAATGAG	420	
27	QY 3540	ATTCCAGTCGAGCGTTAAGCTG-----ATATATATATATGTAATAGGGTG	3585	
28	Db 421	ATTTCAGTCAACGCGTTAAGCTGATATACATATATATATATATATATATATATATATATAT	480	
29	QY 3586	TGATCATAGAAAACCTGTTATGCAATGTTGACTACCTTTGCTTTAAAACCTGCATGAAG	3645	
30	Db 481	TGATCATAGAAAACCTGCTATGCGTTGTATACACCTTTTGTATTCAAAACCTGGATGAAC	540	
31	QY 3646	CTGCAACATA 3655		
32	Db 541	CTGTAACATA 550		
33	RESULT 14			
34	BG592782			
35	LOCUS			
36	DEFINITION	EST491460 csts Solanum tuberosum cDNA clone csts2k4 5' sequence,		
37	FEATURES	529 bp mRNA linear		
38	EST12-APR-2001			

QY 3187 ATACTTAGTGGTGCATTATGAAGGCATC 3216  
|||||  
Db 500 AACTATTAGTGGTGCATCATGAAGACATC 529  
|||||

RESULT 15  
AI896465 421 bp mRNA linear EST 18-MAY-2001  
LOCUS EST265896 tomato callus, TAMU Lycopersicon esculentum cDNA clone  
DEFINITION cLEC15G23, mRNA sequence.  
ACCESSION AI896465  
VERSION AI896465  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 421)  
Alcala, J., Vrebalov, J., White, R., Matern, A.L., Vision, T., Holt, I.E.,  
, Liang, F., Upton, J., Craven, M.B., Bowman, C.L., Ahn, S., Ronning,  
, C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.  
Generation of ESTs from tomato callus tissue  
Unpublished (1999)  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.  
Location/Qualifiers  
1..421  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLEC15G23"  
/clone.lib="tomato callus, TAMU"  
/tissue\_type="callus"  
/dev\_stage="25-40 days old"  
/lab\_host="XLI-Blue MRF"  
/note="vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons  
of seedlings 7-10 days post-germination were excised, cut  
at both ends and placed on MS medium with no selection.  
Mixed callus was harvested at 25 and 40 days and included  
undifferentiated masses. Tomato callus EST Library."  
BASE COUNT 117 a 85 c 101 g 118 t  
ORIGIN

Query Match 11.3%; Score 421; DB 9; Length 421;  
Best Local Similarity 100.0%; Pred. No. 7.4e-97;  
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2464 TCCATGCTTGCATCCTCGTGATTCGGTGTAAAGGCTGTAAATGTCGAGCGCTGCA 2523  
|||||  
Db 1 TCCATGCTTGCATCCTCGTGATTCGGTGTAAAGGCTGTAAATGTCGAGCGCTGCA 60  
|||||

QY 2524 CCACATGGTAGATTGTTGTTATTCCTCAGAAAGAAAAGACCTCATCCGAATGAATG 2583  
|||||  
Db 61 CCACATGGTAGATTGTTGTTATTCCTCAGAAAGAAAAGACCTCATCCGAATGAATG 120  
|||||

QY 2584 TTCTGGGAGTATTGGATGGGATATCTACTTTGTTGGGATCAAGACATGATCCC 2643  
|||||  
Db 121 TTCTGGGAGTATTGGATGGGATATCTACTTTGTTGGGATCAAGACATGATCCC 180  
|||||

QY 2644 GCCAAGGCAAGTCCAGCCGATGGAATATCCTCCAGCACCAGCATACAGTTGGACCATGA 2703  
|||||  
Db 181 GCCAAGGCAAGTCCAGCCGATGGAATATCCTCCAGCACCAGCATACAGTTGGACCATGA 240  
|||||

QY 2704 TGTCAAAATTGAGGAAGTTGAAGAGTACTTCCAACTATATTGTGAATGCAGTTTGGG 2763  
|||||

Db 241 TGTCAAAATTGAGGAAGTTGAAGAGTACTTCCAACTATATTGTGAATGCAGTTTGGG 300  
QY 2764 AATCATAGCAAAATGCCCATGTTCGTATTTCGAGACAGAGAAACCTGATATGCCCATGAGTGA 2823  
|||||  
Db 301 AATCATAGCAAAATGCCCATGTTCGTATTTCGAGACAGAGAAACCTGATATGCCCATGAGTGA 360  
|||||

QY 2824 TCCATGCAAAAAAAGTTGCTGAGCTCTTTTCAATTGCAAGTGGACTTTCCAAAGACTGGTGT 2883  
|||||  
Db 361 TCCATGCAAAAAAAGTTGCTGAGCTCTTTTCAATTGCAAGTGGACTTTCCAAAGACTGGTGT 420  
QY 2884 T 2884  
Db 421 T 421

Search completed: November 5, 2002, 23:19:25  
Job time : 4565 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2002, 23:52:48 ; Search time 148 Seconds  
(without alignments)  
1002.981 Million cell updates/sec

Title: US-09-782-874-2  
Perfect score: 1114  
Sequence: 1 MGKTIQVGFYLLSAEVVK.....RPVNLSSLRQALSHRLVLK 1114

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database : A\_Geneseq\_101002.\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1114	100.0	1114	22 AAU10006	Tomato RNA-directe
2	1114	100.0	1114	22 AAE00897	Tomato C-protein h
3	218	19.6	218	22 AAE00898	Tomato RNA-directe
4	22	2.0	1125	21 AAB28533	Soybean RNA-direct
5	19	1.7	1108	21 AAB28530	Maize RNA-directe
6	17	1.5	17	22 AAU10009	Tomato RNA-directe
7	17	1.5	17	22 AAE00901	Tomato C-protein-s
8	15	1.3	230	21 AAB28534	Maize RNA-directe
9	14	1.3	14	22 AAU10007	Tomato RNA-directe
10	14	1.3	14	22 AAU10008	Tomato RNA-directe

11 1.3 14 22 AAE00899  
12 1.3 14 22 AAE00900  
13 1.3 417 21 AAB28529  
14 1.1 1196 22 AAG67235  
15 1.0 11 22 AAE00903  
16 1.0 13 22 AAU10010  
17 1.0 13 22 AAE00902  
18 1.0 836 21 AAB28532  
19 1.0 1164 21 AAB28531  
20 9 0.8 60 21 AAB28537  
21 8 0.7 93 22 AAC03533  
22 8 0.7 130 21 AAY70941  
23 8 0.7 283 22 AAU18235  
24 8 0.7 327 21 AAY70947  
25 8 0.7 341 20 AAY09307  
26 8 0.7 341 20 AAY09308  
27 8 0.7 477 22 AAU18155  
28 8 0.7 641 23 AAO21854  
29 8 0.7 830 23 ABB93071  
30 8 0.7 1600 22 ABB09417  
31 8 0.7 1600 22 ABB09723  
32 8 0.7 1600 22 ABB12275  
33 8 0.7 1600 22 ABB20004  
34 8 0.7 2168 22 ABB09420  
35 8 0.7 2168 22 ABB09727  
36 8 0.7 2176 22 ABB12287  
37 8 0.7 2176 22 ABB20022  
38 7 0.6 18 22 AAU10011  
39 7 0.6 18 22 ABB43113  
40 7 0.6 18 22 ABB64015  
41 7 0.6 18 22 AAM76838  
42 7 0.6 18 22 AAM35944  
43 7 0.6 18 22 AAM35944  
44 7 0.6 18 22 AAE00904  
45 7 0.6 24 22 ABB44400  
46 7 0.6 24 22 AAM65478  
47 7 0.6 24 22 AAM22018  
48 7 0.6 26 20 AAY18363  
49 7 0.6 31 20 AAY18362  
50 7 0.6 36 21 AAB33978  
51 7 0.6 40 21 AAB25008

#### ALIGNMENTS

##### RESULT 1

AAU10006  
ID AAU10006 standard; Protein; 1114 AA.

XX AC AAU10006;  
XX DT 08-MAY-2002 (first entry)

XX DE Tomato RNA-directed RNA polymerase (RdRP) protein.

XX KW Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;

XX KW cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;

XX KW transgenic plant; transgenic animal; cancer; viral infection;

XX KW immunoprecipitation; immunolocalisation; EC 2.7.7.48; gene therapy.

OS Lycopersicon esculentum.

PN US2001023067-A1.

XX PD 20-SEP-2001.

XX PF 08-FEB-2001; 2001US-0782874.

XX PR 05-MAR-1997; 97US-0811583.

XX PA (WASS/) WASSNEGGER M.

XX PA (RIED/) RIEDEL L.



FT Region 784..797  
 FT /label= P433  
 FT /note= "C-protein-specific antibody for  
 FT immunisation of rabbits"  
 FT 1007..1020  
 FT /label= P431  
 FT /note= "C-protein-specific antibody for  
 FT immunisation of rabbits"  
 FT US6218142-B1.  
 PN 17-APR-2001.  
 XX  
 XX  
 XX 05-MAR-1997; 97US-0811583.  
 XX  
 XX 05-MAR-1997; 97US-0811583.  
 XX (WASS/) WASENEGGER M.  
 XX (RIED/) RIEDEL L.  
 XX  
 XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;  
 XX WPI; 2001-289830/30.  
 XX N-PSDB: AAD04370.  
 DR  
 DR New nucleic acid molecules encoding polypeptides with RNA-directed RNA  
 PT polymerase enzymatic activity, useful in modulating gene expression in  
 PT plants, humans and animals, as well as in plant cell/tissue cultures or  
 PT plant breeding -  
 XX  
 XX Claim 1; Column 33-40; 31pp; English.  
 XX  
 CC The present sequence is a tomato C-protein having RNA-directed RNA  
 CC polymerase (RdRP) activity, this protein is capable of RNA-directed RNA  
 CC synthesis, thus using RNA as a template for synthesizing complementary  
 CC RNA molecules. RdRP nucleic acid is useful for modulating gene expression  
 CC in plants, humans and animals. This may lead to various physiological,  
 CC developmental and/or morphological changes. Transgenic plants containing  
 CC RdRP nucleic acid is especially useful in plant cell or tissue cultures  
 CC and in plant breeding. RdRP is useful in gene therapy, particularly for  
 CC treating a disease that is caused by the undesirable expression or  
 CC overexpression of a gene.  
 XX  
 XX Sequence 1114 AA;  
 SQ  
 Query Match 100.0%; Score 1114; DB 22; Length 1114;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MGKTIQVGFYLLSAEVVKSFLKYTGTVCALEVKQSGSGRAFAKVQFADNISADK 60  
 DB 1 MGKTIQVGFYLLSAEVVKSFLKYTGTVCALEVKQSGSGRAFAKVQFADNISADK 60  
 QY 61 IITLANNRLYTGSSYLKAWEMKTDIVQLRAYVDQMDGTTLNFQCISDDKFAVLGSTEVS 120  
 DB 61 IITLANNRLYTGSSYLKAWEMKTDIVQLRAYVDQMDGTTLNFQCISDDKFAVLGSTEVS 120  
 QY 121 IQFGIGLKKFFFLSSGSADYKQLQSYENIQVWLHRLPYGQNAQFLLIQLFGAPRIYKRL 180  
 DB 121 IQFGIGLKKFFFLSSGSADYKQLQSYENIQVWLHRLPYGQNAQFLLIQLFGAPRIYKRL 180  
 QY 181 ENSCYSFKEPDPDQWVTTDFPPSWIGLSSSLCQLQPRGVRLENFESFFHYAERENNI 240  
 DB 181 ENSCYSFKEPDPDQWVTTDFPPSWIGLSSSLCQLQPRGVRLENFESFFHYAERENNI 240  
 QY 241 TIQTGTFEFVSKSALVENVOPPEGISIPYKILFKISSLVQHGCIPGALNVYFFRLVDP 300  
 DB 241 TIQTGTFEFVSKSALVENVOPPEGISIPYKILFKISSLVQHGCIPGALNVYFFRLVDP 300  
 QY 301 RRRNVACIEHALEKLYIYKECCYDPVRWLTEQYDGYLKGROPKPSPTITLDDGLVYVRRV 360  
 DB 301 RRRNVACIEHALEKLYIYKECCYDPVRWLTEQYDGYLKGROPKPSPTITLDDGLVYVRRV 360

QY 361 LVTPCKVYFCGPEVNVSNRVLRYNSIEDINFLRVSFVDEEWKLYSTDLLPKASTGSGVR 420  
 DB 361 LVTPCKVYFCGPEVNVSNRVLRYNSIEDINFLRVSFVDEEWKLYSTDLLPKASTGSGVR 420  
 QY 421 TNIIYERILSTLRKGFVIGDKKFFELAFSSQLRDNSVWMEFASRPGLTANDIRAWMGDFSQ 480  
 DB 421 TNIIYERILSTLRKGFVIGDKKFFELAFSSQLRDNSVWMEFASRPGLTANDIRAWMGDFSQ 480  
 QY 481 IKNVAKYAARLGQSGFSSRRETTLSVLRHEIEVIPDVKVHGTSYVFSGIGISGDFAHRA 540  
 DB 481 IKNVAKYAARLGQSGFSSRRETTLSVLRHEIEVIPDVKVHGTSYVFSGIGISGDFAHRA 540  
 QY 541 SKCGLQYTPSAFQIRYGGYKGVGVDPDSSMKLSLRKSMKSYESDNKIKLDVLGWSKYQPC 600  
 DB 541 SKCGLQYTPSAFQIRYGGYKGVGVDPDSSMKLSLRKSMKSYESDNKIKLDVLGWSKYQPC 600  
 QY 601 YLNRQLITLLSTLGVKDEVLEQKQKAVDQDAILHDSLKAQAEALMSGENTNLIKAM 660  
 DB 601 YLNRQLITLLSTLGVKDEVLEQKQKAVDQDAILHDSLKAQAEALMSGENTNLIKAM 660  
 QY 661 LNCGYKPDABEFTLSMMLQTPRASKLLDLRTRSRIFPNGRTNMGCDESTLEYGQVFFVQ 720  
 DB 661 LNCGYKPDABEFTLSMMLQTPRASKLLDLRTRSRIFPNGRTNMGCDESTLEYGQVFFVQ 720  
 QY 721 FTGAGHGEFSDDLHPFNNSRSTNSNFIKGNVYVAKNPCLHPGDIRVLKAVNVRALHHMV 780  
 DB 721 FTGAGHGEFSDDLHPFNNSRSTNSNFIKGNVYVAKNPCLHPGDIRVLKAVNVRALHHMV 780  
 QY 781 DCVVFQKGRPHNECSGSDLDGDIYFVCWQDMIPPRQVQPMYPPAPSIQLDHDTV 840  
 DB 781 DCVVFQKGRPHNECSGSDLDGDIYFVCWQDMIPPRQVQPMYPPAPSIQLDHDTV 840  
 QY 841 EEVEEFTNIVNDSGLIIANAHVVFADREPDMAMSDPCKKLAELFSIAVDPEKTVPAE 900  
 DB 841 EEVEEFTNIVNDSGLIIANAHVVFADREPDMAMSDPCKKLAELFSIAVDPEKTVPAE 900  
 QY 901 IPSQLRPEKYPFMDKPKDTSYISERVIGKLFKVKDKAPQASSIATFTRDVARRSYDAD 960  
 DB 901 IPSQLRPEKYPFMDKPKDTSYISERVIGKLFKVKDKAPQASSIATFTRDVARRSYDAD 960  
 QY 961 MEVDGFEDYIDEAFDYKTEYDNKLGNDYXGKTEAEILSGGIMKASKTFDRRKDAEAI 1020  
 DB 961 MEVDGFEDYIDEAFDYKTEYDNKLGNDYXGKTEAEILSGGIMKASKTFDRRKDAEAI 1020  
 QY 1021 SVAVRALRKEARAWFKRRNDIDDMPLKASAWYHVTYHTYWGYNQGLKRAHFI SFPWCY 1080  
 DB 1021 SVAVRALRKEARAWFKRRNDIDDMPLKASAWYHVTYHTYWGYNQGLKRAHFI SFPWCY 1080  
 QY 1081 YDQLIQIKKDKARNRPVNLSSLRQAQLSHRLVLK 1114  
 DB 1081 YDQLIQIKKDKARNRPVNLSSLRQAQLSHRLVLK 1114  
 RESULT 3  
 AAE00898  
 ID AAE00898 standard; Protein; 218 AA.  
 XX  
 XX AAE00898;  
 XX AC  
 XX XX  
 XX DT  
 XX 04-JUL-2001 (first entry)  
 XX  
 XX Tomato RNA-directed RNA polymerase 800 (RdRP800) protein.  
 XX  
 XX Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;  
 KW transgenic plant; tissue culture; plant breeding; therapy; C-protein.  
 XX  
 XX Lycopersicon esculentum.  
 XX  
 XX US6218142-B1.  
 XX  
 XX 17-APR-2001.  
 XX  
 XX 05-MAR-1997; 97US-0811583.  
 XX  
 XX PF

```

XX 05-MAR-1997; 97US-0811583.
PR
XX (WASS/) WASSENEGGER M.
PA (RIED/) RIEDEL L.
XX
XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX WPI; 2001-289830/30.
DR
XX
XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA
PT polymerase enzymatic activity, useful in modulating gene expression in
PT plants, humans and animals, as well as in plant cell/tissue cultures or
PT plant breeding -
XX
XX Example 3; Column 39-42; 31pp; English.
PS
XX
XX The present sequence is tomato RNA-directed RNA polymerase 800
CC (RdRP00) protein which corresponds to position 700-917 of C-protein.
CC C-protein is capable of RNA-directed RNA synthesis, thus using RNA as a
CC template for synthesizing complementary RNA molecules. RdRP nucleic acid
CC is useful for modulating gene expression in plants, humans and animals.
CC This may lead to various physiological, developmental and/or
CC morphological changes. Transgenic plants containing RdRP nucleic acid is
CC especially useful in plant cell or tissue cultures and in plant breeding.
CC RdRP is useful in gene therapy, particularly for treating a disease that
CC is caused by the undesirable expression or overexpression of a gene.
XX
XX Sequence 218 AA;
SQ
Query Match 19.6%; Score 218; DB 22; Length 218;
Best Local Similarity 100.0%; Pred. No. 4.3e-224; Indels 0; Gaps 0;
Matches 218; Conservative 0; Mismatches 0;
QY 700 RTMGCLDESRTLEYGVQVFTGAGHGFSDDLHPFNNSRSTNSFILKGNVYVAKNPC 759
DB 1 RTMGCLDESRTLEYGVQVFTGAGHGFSDDLHPFNNSRSTNSFILKGNVYVAKNPC 60
QY 760 LHPGDIRVLKAVNRALHHMVDVVFQKGRPHNECSGSLDGDIFYFCWDQDMTPPR 819
DB 61 LHPGDIRVLKAVNRALHHMVDVVFQKGRPHNECSGSLDGDIFYFCWDQDMTPPR 120
QY 820 VQQMEYPPAPSIQLDHDVTIEVEEYFTNYIVNDSLGIANAHVVFADREPDMAMSDPC 879
DB 121 VQQMEYPPAPSIQLDHDVTIEVEEYFTNYIVNDSLGIANAHVVFADREPDMAMSDPC 180
QY 880 KKLAEFSIAVDFPKTGVPAIPISQLRPKEYPDMFKP 917
DB 181 KKLAEFSIAVDFPKTGVPAIPISQLRPKEYPDMFKP 218
RESULT 4
AAB28533
ID AAB28533 standard; Protein; 1125 AA.
XX
XX AAB28533;
XX
XX 07-FEB-2001 (first entry)
XX
XX Soybean RNA-directed RNA polymerase, SEQ ID NO: 12.
DE
XX
XX Soybean; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
KW plant virus resistance; plant breeding.
XX
XX Glycine max.
OS
XX WO2000060097-A1.
PN
XX 12-OCT-2000.
PD
XX
XX 06-APR-2000; 2000WO-US09105.
PS
XX
XX 07-APR-1999; 99US-0128094.
XX
XX The present sequence is part of a plant RNA-directed RNA polymerase
CC

```

```

XX (DUPO ) DU PONT DE NEMOURS & CO E I.
PA (PION-) PIONEER HI-BRED INT INC.
XX
XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;
XX WPI; 2000-679376/66.
DR N-PSDB; AAC63742.
XX
XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT for controlling gene expression and providing mechanisms to engineer
PT plant virus resistance -
XX
XX Claim 10; Page 50-54; 62pp; English.
PS
XX
XX The present sequence is part of a plant RNA-directed RNA polymerase
CC protein. Polynucleotides encoding RNA-directed RNA polymerases were
CC isolated from plant cDNA libraries. They are useful as probes for
CC genetically and physically mapping genes, and as markers for
CC traits linked to those genes. They are useful for controlling gene
CC expression and provide mechanisms to engineer plant virus resistance.
CC They are also useful for plant breeding to develop lines with desired
CC phenotypes.
XX
XX Sequence 1125 AA;
SQ
Query Match 2.0%; Score 22; DB 21; Length 1125;
Best Local Similarity 100.0%; Pred. No. 9.9e-14; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;
QY 791 RHPNECSGSLDGDIFYFCWD 812
DB 801 RHPNECSGSLDGDIFYFCWD 822
RESULT 5
AAB28530
ID AAB28530 standard; Protein; 1108 AA.
XX
XX AAB28530;
XX
XX 07-FEB-2001 (first entry)
XX
XX Maize RNA-directed RNA polymerase, SEQ ID NO: 6.
DE
XX
XX Maize; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
KW plant virus resistance; plant breeding.
XX
XX Zea mays.
OS
XX WO2000060097-A1.
PN
XX 12-OCT-2000.
PD
XX
XX 06-APR-2000; 2000WO-US09105.
PF
XX
XX 07-APR-1999; 99US-0128094.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
PA (PION-) PIONEER HI-BRED INT INC.
XX
XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;
XX WPI; 2000-679376/66.
DR N-PSDB; AAC63739.
XX
XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT for controlling gene expression and providing mechanisms to engineer
PT plant virus resistance -
XX
XX Claim 10; Page 36-40; 62pp; English.
PS
XX
XX The present sequence is part of a plant RNA-directed RNA polymerase
CC

```

CC protein. Polynucleotides encoding RNA-directed RNA polymerases were  
 CC isolated from plant cDNA libraries. They are useful as probes for  
 CC genetically and physically mapping genes, and as markers for  
 CC traits linked to those genes. They are useful for controlling gene  
 CC expression and provide mechanisms to engineer plant virus resistance.  
 CC They are also useful for plant breeding to develop lines with desired  
 CC phenotypes.

XX Sequence 1108 AA;

Query Match 1.7%; Score 19; DB 21; Length 1108;

Best Local Similarity 100.0%; Pred. No. 1.6e-10; Mismatches 0; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 791 RPHNECSGSLDGDYFV 809

|||||

Db 794 RPHNECSGSLDGDYFV 812

RESULT 6

AAU10009  
 ID AAU10009 standard; Peptide; 17 AA.

XX AC AAU10009;

XX DT 08-MAY-2002 (first entry)

XX DE Tomato RNA-directed RNA polymerase (RdRP) immunogenic peptide P432.

XX KW Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;  
 KW cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;  
 KW transgenic plant; transgenic animal; cancer; viral infection;  
 KW immunoprecipitation; immunolocalisation; P432; gene therapy.

XX OS Lycopersicon esculentum.

XX US2001023067-A1.

XX PD 20-SEP-2001.

XX PF 08-FEB-2001; 2001US-0782874.

XX PR 05-MAR-1997; 97US-0811583.

XX PA (WASS/) WASSENEGGER M.  
 XX PA (RIED/) RIEDEL L.

XX PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;

XX DR WPI; 2001-595798/67.

XX PT New nucleic acid molecule encoding a polypeptide having the enzymatic  
 PT activity of RNA-directed RNA polymerase, for modulating gene expression  
 PT and treating cancer and virus infection in human and animals

XX PS Example 7; Page 14; 34pp; English.

XX This sequence represents the tomato RNA-directed RNA polymerase (RdRP)  
 CC immunogenic peptide P432 used to immunise rabbits and produce an RdRP-  
 CC specific antibody. The invention comprises the nucleic acid and  
 CC protein sequences of RdRP, the protein of the invention can catalyse in  
 CC vitro transcription of short single stranded RNAs into DNA molecules,  
 CC this transcription can be either primed by RNA or DNA oligonucleotides  
 CC or be unprimed. The protein may have cytosolic or virucide activities.  
 CC The sequences of the invention may be used in gene therapy or as an RNA  
 CC directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template  
 CC nucleic acid molecule derived from a nucleic acid molecule which causes  
 CC a disease are useful for treating a disease caused by the undesired  
 CC expression or overexpression of a nucleic acid molecule in a human, rat  
 CC or mouse, by administering the molecules. This system can be used in the  
 CC preparation of a pharmaceutical composition and for inhibiting  
 CC expression of any desired gene by transferring the RdRP system to  
 CC organisms that either lack a comparable mechanism or do not sufficiently

CC express their own RdRP. An antibody or an antagonist or inhibitor to the  
 CC protein are useful for inhibiting RNA directed RNA synthesis and for  
 CC ensuring stable heterologous, gene expression in transgenic organisms.  
 CC The sequence is useful for probes and/or for the control of gene  
 CC expression, as primers for amplification of nucleic acid molecules and  
 CC as tools for the detection of expression of the cDNA molecules.

CC Additionally, nucleotide and protein sequences are useful for  
 CC suppression of undesired gene expression in humans and animals. The RdRP  
 CC is useful as a therapeutic agent for the control of cancer and virus  
 CC infection in humans and animals and the antibody is useful for  
 CC immunoprecipitation or immunolocalisation of the protein, identification  
 CC of polypeptides interacting with it and screening expression libraries.

XX Sequence 17 AA;

Query Match 1.5%; Score 17; DB 22; Length 17;

Best Local Similarity 100.0%; Pred. No. 4.1e-10;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 331 EYDGYLKGROPPKSPS 347

|||||

Db 1 EYDGYLKGROPPKSPS 17

RESULT 7

AAE00901  
 ID AAE00901 standard; peptide; 17 AA.

XX AC AAE00901;

XX DT 04-JUL-2001 (first entry)

XX DE Tomato C-protein-specific antibody, P432 for immunisation of rabbits.

XX KW Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;  
 KW transgenic plant; tissue culture; plant breeding; therapy; C-protein;  
 KW antibody; immunisation.

XX OS Lycopersicon esculentum.

XX PN US6218142-B1.

XX PD 17-APR-2001.

XX PF 05-MAR-1997; 97US-0811583.

XX PR 05-MAR-1997; 97US-0811583.

XX PA (WASS/) WASSENEGGER M.  
 XX PA (RIED/) RIEDEL L.

XX PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;

XX DR WPI; 2001-289830/30.

XX PT New nucleic acid molecules encoding polypeptides with RNA-directed RNA  
 PT polymerase enzymatic activity, useful in modulating gene expression in  
 PT plants, humans and animals, as well as in plant cell/tissue cultures or  
 PT plant breeding

XX PS Example 7; Column 24; 31pp; English.

XX The present sequence is a tomato C-protein-specific antibody, P432 chosen  
 CC for immunisation of rabbits. C-protein having RNA-directed RNA polymerase  
 CC (RdRP) activity is capable of RNA-directed RNA synthesis, thus using RNA  
 CC as a template for synthesising complementary RNA molecules. RdRP nucleic  
 CC acid is useful for modulating gene expression in plants, humans and  
 CC animals. This may lead to various physiological, developmental and/or  
 CC morphological changes. Transgenic plants containing RdRP nucleic acid is  
 CC especially useful in plant cell or tissue cultures and in plant breeding.  
 CC RdRP is useful in gene therapy, particularly for treating a disease that  
 CC is caused by the undesirable expression or overexpression of a gene.





RESULT	10
AAU10008	
ID	AAU10008 standard; Peptide; 14 AA.
XX	
AC	AAU10008;
XX	
DT	08-MAY-2002 (first entry)
XX	
Tomato RNA-directed RNA polymerase (RdRP) immunogenic peptide P431.	
XX	
DE	Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
XX	Cytostatic; virucide; RNA synthesis inhibitor; antibody; immunogen;
KW	transgenic plant; transgenic animal; cancer; viral infection;
KW	immunoprecipitation; immunolocalisation; P431; gene therapy.
XX	
Lycopodium esculentum.	
OS	
US2001023067-A1.	
XX	
PD	20-SEP-2001.
XX	
Pf	08-FEB-2001; 2001US-0782874 .
XX	
PPR	05-MAR-1997; 97US-0811583.
XX	
(WASS/) WASSENEGGER M.	
PA	(RIED/) RIEDEL L.
XX	
Wassenegger M, Riedel L, Schielbel W, Sanger HL;	
PI	
DR	WPI; 2001-595798/67.
XX	
New nucleic acid molecule encoding a polypeptide having the enzymatic	
PT	activity of RNA-directed RNA polymerase, for modulating gene expression
PT	and treating cancer and virus infection in human and animals -
PT	
XX	Example 7; Page 14; 34pp; English.
XX	
This sequence represents the tomato RNA-directed RNA polymerase (RdRP)-	
CC	immunogenic peptide P431 used to immunise rabbits and produce an RdRp-
CC	specific antibody. The invention comprises the nucleic acid and
CC	protein sequences of RdRP, the protein of the invention can catalyse in
CC	viro transcription of short single stranded RNAs into DNA molecules,
CC	this transcription can be either primed by RNA or DNA oligonucleotides,
CC	or be unprimed. The protein may have cytotstatic or virucide activities.
CC	The sequences of the invention may be used in gene therapy or as an RNA
CC	directed RNA synthesis inhibitor. The RdRp cDNA sequence and a template
CC	nucleic acid molecule derived from a nucleic acid molecule which causes
CC	a disease are useful for treating a disease caused by the undesired
CC	expression or overexpression of a nucleic acid molecule in a human, rat
CC	or mouse, by administering the molecules. This system can be used in the
CC	preparation of a pharmaceutical composition and for inhibiting
CC	expression of any desired gene by transferring the RdRp system to
CC	organisms that either lack a comparable mechanism or do not sufficiently
CC	express their own RdRp. An antibody or an antagonist or inhibitor to the
CC	protein are useful for inhibiting RNA directed RNA synthesis and for
CC	ensuring stable heterologous, gene expression in transgenic organisms.
CC	The sequence is useful for probes and/or for the control of gene
CC	expression, as primers for amplification of nucleic acid molecules and
CC	as tools for the detection of expression of the cDNA molecules.
CC	Additionally, nucleotide and protein sequences are useful for
CC	suppression of undesired gene expression in humans and animals. The RdRp
CC	is useful as a therapeutic agent for the control of cancer and virus
CC	infection in humans and animals and the antibody is useful for
CC	immunoprecipitation or immunolocalisation of the protein, identification
CC	of polypeptides interacting with it and screening expression libraries.
XX	
SQ	Sequence 14 AA;

Query Match            1.3%; Score 14; DB 22; Length 14;  
Best Local Similarity    100.0%; Pred. No. 5.6e-07;  
Matches 14; Conservative      0; Mismatches     0; Indels       0; Gaps          0;

DE Tomato C-protein-specific antibody, P431 for immunisation of rabbits.  
XX  
KW Tomato: gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;  
KW transgenic plant; tissue culture; plant breeding; therapy; C-protein;  
KW antibody; immunisation.  
XX  
OS Lycopersicon esculentum.  
XX  
PN US6218142-B1.  
XX  
PD 17-APR-2001.  
XX  
XX 05-MAR-1997; 9705-0811583.  
PF  
XX  
PR 05-MAR-1997; 9705-0811583.  
XX  
XX (WASS/) WASSENEGGER M.  
PA (RIED/) RIEDEL L.  
XX  
XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;  
XX WPI; 2001-289830/30.  
DR  
XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA  
XX polymerase enzymatic activity, useful in modulating gene expression in  
XX plants, humans and animals, as well as in plant cell/tissue cultures or  
XX plant breeding -  
XX  
XX Example 7; Column 24; 31pp; English.  
XX  
XX The present sequence is a tomato C-protein-specific antibody, P431 chosen  
XX for immunisation of rabbits. C-protein having RNA-directed RNA polymerase  
XX (RdRP) activity is capable of RNA-directed RNA synthesis, thus using RNA  
XX as a template for synthesising complementary RNA molecules. RdRP nucleic  
XX acid is useful for modulating gene expression in plants, humans and  
XX animals. This may lead to various physiological, developmental and/or  
XX morphological changes. Transgenic plants containing RdRP nucleic acid is  
XX especially useful in plant cell or tissue cultures and in plant breeding.  
XX RdRP is useful in gene therapy, particularly for treating a disease that  
XX is caused by the undesirable expression or overexpression of a gene.  
XX  
XX Sequence 14 AA;  
SQ  
  
Query Match 1.3%; Score 14; DB 22; Length 14;  
Best Local Similarity 100.0%; Pred. No. 5.6e-07;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1007 ASKTFDRRKDAEAI 1020  
DB 1 ASKTFDRRKDAEAI 14  
  
RESULT 13  
AAB28529  
ID AAB28529 standard; Protein; 417 AA.  
XX  
AC AAB28529;  
XX  
XX 07-FEB-2001 (first entry)  
XX  
XX Maize partial RNA-directed RNA polymerase, SEQ ID NO: 4.  
XX  
XX Maize; plant; RNA-directed RNA polymerase; gene mapping; gene marker;  
XX plant virus resistance; plant breeding.  
XX  
XX Zea mays.  
XX  
XX WO2000060097-A1.  
XX  
XX 12-OCT-2000.  
XX  
XX 06-APR-2000; 2000WO-US09105.  
XX  
XX

PR 07-APR-1999; 99US-0128094.  
XX  
XX (DUPO) DU PONT DE NEMOURS & CO E I.  
PA (PION-) PIONEER HI-BRED INT INC.  
XX  
XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;  
PI WPI; 2000-679376/66.  
XX N-PSDB; AAC63738.  
DR  
XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful  
XX for controlling gene expression and providing mechanisms to engineer  
XX plant virus resistance -  
XX  
XX Claim 10; Page 34-35; 62pp; English.  
XX  
XX The present sequence is part of a plant RNA-directed RNA polymerase  
XX protein. Polynucleotides encoding RNA-directed RNA polymerases were  
XX isolated from plant cDNA libraries. They are useful as probes for  
XX genetically and physically mapping genes, and as markers for  
XX traits linked to those genes. They are useful for controlling gene  
XX expression and provide mechanisms to engineer plant virus resistance.  
XX They are also useful for plant breeding to develop lines with desired  
XX phenotypes.  
XX  
XX Sequence 417 AA;  
SQ  
  
Query Match 1.3%; Score 14; DB 21; Length 417;  
Best Local Similarity 100.0%; Pred. No. 1.4e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 549 PSAPQIRYGGYKGV 562  
DB 62 PSAPQIRYGGYKGV 75  
  
RESULT 14  
AAG67235  
ID AAG67235 standard; Protein; 1196 AA.  
XX  
AC AAG67235;  
XX  
XX 13-NOV-2001 (first entry)  
XX  
XX Amino acid sequence of a plant SGS2 polypeptide.  
DE  
XX  
XX SGS2: RNA-dependent RNA polymerase; transgene silencing;  
KW transgene stability; crop plant; viral resistance.  
KW  
XX Arabidopsis thaliana.  
OS  
XX FR2804128-A1.  
XX  
XX 27-JUL-2001.  
PD  
XX  
XX 26-JAN-2000; 2000FR-0001007.  
PF  
XX  
XX 26-JAN-2000; 2000FR-0001007.  
PR  
XX (RHOB-) RHOBIO.  
PA (INRG) INST NAT RECH AGRONOMIQUE.  
XX  
XX Beclin C, Elmayer T, Mourrain P, Vaucheret H;  
PI WPI; 2001-543303/61.  
XX N-PSDB; AAH77704.  
DR  
XX  
XX A new plant SGS2 gene involved in encoding an RNA-dependent RNA  
XX polymerase and in transgene silencing, increases transgene stability  
XX and expression in transgenic plants when it is inactivated -  
XX  
XX Claim 24; Fig 1; 46pp; French.  
XX

CC The present sequence represents a plant SGS2 polypeptide. SGS2 is an  
 CC RNA-dependent RNA polymerase and is involved in transgene silencing.  
 CC Inactivation of SGS2 is used to increase transgene stability and  
 CC expression in plants, particularly crop plants, especially maize, corn,  
 CC barley, sorghum, soy, sugar cane, beet, tobacco or cotton plants.  
 CC Overexpression of SGS2 can be used to increase resistance to viral  
 CC infection in plants.

XX SQ Sequence 1196 AA;

Query Match 1.1%; Score 12; DB 22; Length 1196;  
 Best Local Similarity 100.0%; Pred. No. 0.0052;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1047 KASAWHYTYHP 1058  
 DB 1124 KASAWHYTYHP 1135  
 |||||

# RESULT 15

AAE00903  
 ID AAE00903 standard; peptide; 11 AA.

AC AAE00903;

DT 04-JUL-2001 (first entry)

XX Peptide #1 obtained by micro-sequencing RdRP protein.

DE Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;  
 KW transgenic plant; tissue culture; plant breeding; therapy; C-protein.

XX Lycopersicon esculentum.

OS US6218142-B1.

PN 17-APR-2001.

XX 05-MAR-1997; 97US-0811583.

PF 05-MAR-1997; 97US-0811583.

PR (WASS/) WASENEGGER M.

XX (RIED/) RIEDEL L.

PI Wassenegeger M, Riedel L, Schiebel W, Sanger HL;

XX WPI; 2001-289830/30.

DR N-PSDB: AAD04381.

XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA  
 PT polymerase enzymatic activity, useful in modulating gene expression in  
 PT plants, humans and animals, as well as in plant cell/tissue cultures or  
 PT plant breeding

XX Example 3; Fig 6; 31pp; English.

CC The present sequence is a peptide encoded by an oligonucleotide which is  
 CC used to design RNA-directed RNA polymerase (RdRP)-specific PCR primers.  
 CC These PCR primers are used to amplify a cDNA encoding tomato C-protein  
 CC having RdRP activity. The peptide is obtained by micro-sequencing RdRP  
 CC protein. C-protein is capable of RNA-directed RNA synthesis, thus using  
 CC RNA as a template for synthesising complementary RNA molecules. RdRP  
 CC nucleic acid is useful for modulating gene expression in plants, humans  
 CC and animals. This may lead to various physiological, developmental and/or  
 CC morphological changes. Transgenic plants containing RdRP nucleic acid is  
 CC especially useful in plant cell or tissue cultures and in plant breeding.  
 CC RdRP is useful in gene therapy, particularly for treating a disease that  
 CC is caused by the undesirable expression or overexpression of a gene.

XX SQ Sequence 11 AA;

Query Match 1.0%; Score 11; DB 22; Length 11;

Best Local Similarity 100.0%; Pred. No. 0.00072;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 909 EYPDFMDKPKD 919  
 DB 1 EYPDFMDKPKD 11  
 |||||

# RESULT 16

AAU10010  
 ID AAU10010 standard; Peptide; 13 AA.

AC AAU10010;

DT 08-MAY-2002 (first entry)

XX Tomato RNA-directed RNA polymerase (RdRP) immunogenic peptide P433.

DE Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;  
 KW cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;  
 KW transgenic plant; transgenic animal; cancer; viral infection;  
 KW immunoprecipitation; immunolocalisation; P433; gene therapy.

XX Lycopersicon esculentum.

OS US2001023067-A1.

PN 20-SEP-2001.

XX 08-FEB-2001; 2001US-0782874.

PR 05-MAR-1997; 97US-0811583.

XX (WASS/) WASENEGGER M.

PA (RIED/) RIEDEL L.

PI Wassenegeger M, Riedel L, Schiebel W, Sanger HL;

XX WPI; 2001-595798/67.

XX New nucleic acid molecule encoding a polypeptide having the enzymatic  
 PT activity of RNA-directed RNA polymerase, for modulating gene expression  
 PT and treating cancer and virus infection in human and animals

XX Example 7; Page 14; 34pp; English.

CC This sequence represents the tomato RNA-directed RNA polymerase (RdRP)  
 CC immunogenic peptide P433 used to immunise rabbits and produce an RdRP-  
 CC specific antibody. The invention comprises the nucleic acid and  
 CC protein sequences of RdRP, the protein of the invention can catalyse in  
 CC vitro transcription of short single stranded RNAs into DNA molecules,  
 CC this transcription can be either primed by RNA or DNA oligonucleotides,  
 CC or be unprimed. The protein may have cytosolic or virucide activities.  
 CC The sequences of the invention may be used in gene therapy or as an RNA  
 CC directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template  
 CC nucleic acid molecule derived from a nucleic acid molecule which causes  
 CC a disease are useful for treating a disease caused by the undesired  
 CC expression or overexpression of a nucleic acid molecule in a human, rat  
 CC or mouse, by administering the molecules. This system can be used in the  
 CC preparation of a pharmaceutical composition and for inhibiting  
 CC expression of any desired gene by transferring the RdRP system to  
 CC organisms that either lack a comparable mechanism or do not sufficiently  
 CC express their own RdRP. An antibody or an antagonist or inhibitor to the  
 CC protein are useful for inhibiting RNA directed RNA synthesis and for  
 CC ensuring stable heterologous, gene expression in transgenic organisms.  
 CC The sequence is useful for probes and/or for the control of gene  
 CC expression, as primers for amplification of nucleic acid molecules and  
 CC as tools for the detection of expression of the cDNA molecules.

CC Additionally, nucleotide and protein sequences are useful for  
 CC suppression of undesired gene expression in humans and animals. The RdRP  
 CC is useful as a therapeutic agent for the control of cancer and virus  
 CC infection in humans and animals and the antibody is useful for  
 CC immunoprecipitation or immunolocalisation of the protein, identification

CC of polypeptides interacting with it and screening expression libraries.

```
XX
SQ Sequence 13 AA;
Query Match 0.9%; Score 10; DB 22; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.0099;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 784 VFPOKGRPH 793
|
Db 1 VFPOKGRPH 10
|

RESULT 17
AAE00902
ID AAE00902 standard; peptide; 13 AA.
XX
AC AAE00902;
XX
DT 04-JUL-2001 (first entry)
XX
XX Tomato C-protein-specific antibody, P433 for immunisation of rabbits.
DE
KW Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
KW transgenic plant; tissue culture; plant breeding; therapy; C-protein;
KW antibody; immunisation.
XX
OS Lycopersicon esculentum.
XX
PN US6218142-B1.
XX
PD 17-APR-2001.
XX
PF 05-MAR-1997; 97US-0811583.
XX
PR 05-MAR-1997; 97US-0811583.
XX
PA (WASS/) WASSENEGGER M.
PA (RIED/) RIEDEL L.
XX
XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX
XX WPI; 2001-289830/30.
XX
XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA
XX polymerase enzymatic activity, useful in modulating gene expression in
XX plants, humans and animals, as well as in plant cell/tissue cultures or
XX plant breeding
XX
XX Example 7; Column 24; 3lpp; English.
XX
XX The present sequence is a tomato C-protein-specific antibody, P433 chosen
XX for immunisation of rabbits. C-protein having RNA-directed RNA polymerase
XX (RdRP) activity is capable of RNA-directed RNA synthesis, thus using RNA
XX as a template for synthesising complementary RNA molecules. RdRP nucleic
XX acid is useful for modulating gene expression in plants, humans and
XX animals. This may lead to various physiological, developmental and/or
XX morphological changes. Transgenic plants containing RdRP nucleic acid is
XX especially useful in plant cell or tissue cultures and in plant breeding.
XX RdRp is useful in gene therapy, particularly for treating a disease that
XX is caused by the undesirable expression or overexpression of a gene.
XX
SQ Sequence 13 AA;
Query Match 0.9%; Score 10; DB 22; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.0099;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 784 VFPOKGRPH 793
|
Db 1 VFPOKGRPH 10
|

RESULT 18
AAB28532
ID AAB28532 standard; Protein; 836 AA.
XX
AC AAB28532;
XX
DT 07-FEB-2001 (first entry)
XX
DE Rice RNA-directed RNA polymerase, SEQ ID NO: 10.
XX
KW Rice; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
KW plant virus resistance; plant breeding.
XX
OS Oryza sativa.
XX
PN WO200060097-A1.
XX
PD 12-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US09105.
XX
PR 07-APR-1999; 99US-0128094.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;
XX
XX WPI; 2000-679376/66.
XX
XX N-PSDB; AAC63741.
XX
XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
XX for controlling gene expression and providing mechanisms to engineer
XX plant virus resistance
XX
XX Claim 10; Page 46-49; 62pp; English.
XX
XX The present sequence is part of a plant RNA-directed RNA polymerase
XX protein. Polynucleotides encoding RNA-directed RNA polymerases were
XX isolated from plant cDNA libraries. They are useful as probes for
XX genetically and physically mapping genes, and as markers for
XX traits linked to those genes. They are useful for controlling gene
XX expression and provide mechanisms to engineer plant virus resistance.
XX They are also useful for plant breeding to develop lines with desired
XX phenotypes.
XX
XX Sequence 836 AA;
Query Match 0.9%; Score 10; DB 21; Length 836;
Best Local Similarity 100.0%; Pred. No. 0.51;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 755 AKNPCLHPGD 764
|
Db 459 AKNPCLHPGD 468
|

RESULT 19
AAB28531
ID AAB28531 standard; Protein; 1164 AA.
XX
AC AAB28531;
XX
DT 07-FEB-2001 (first entry)
XX
DE Maize RNA-directed RNA polymerase, SEQ ID NO: 8.
XX
KW Maize; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
KW plant virus resistance; plant breeding.
XX
OS Zea mays.
XX
PN WO200060097-A1.
```

XX 12-OCT-2000.  
 XX 06-APR-2000; 2000WO-US09105.  
 XX 07-APR-1999; 99US-0128094.  
 XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;  
 XX WPI; 2000-679376/66.  
 DR N-PSDB; AAC63740.  
 XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful  
 PT for controlling gene expression and providing mechanisms to engineer  
 PT plant virus resistance -  
 XX Claim 10; Page 41-45; 62pp; English.  
 XX The present sequence is part of a plant RNA-directed RNA polymerase  
 CC protein. Polynucleotides encoding RNA-directed RNA polymerases were  
 CC isolated from plant cDNA libraries. They are useful as probes for  
 CC genetically and physically mapping genes, and as markers for  
 CC traits linked to those genes. They are useful for controlling gene  
 CC expression and provide mechanisms to engineer plant virus resistance.  
 CC They are also useful for plant breeding to develop lines with desired  
 CC phenotypes.  
 XX Sequence 1164 AA;  
 SQ Query Match 0.9%; Score 10; DB 21; Length 1164;  
 Best Local Similarity 100.0%; Pred. No. 0.7;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 755 AKNPLCHPGD 764  
 DB 787 AKNPLCHPGD 796  
 RESULT 20  
 AAB28537  
 ID AAB28537 standard; Protein; 60 AA.  
 AC  
 AC AAB28537;  
 DT 07-FEB-2001 (first entry)  
 XX Soybean RNA-directed RNA polymerase, SEQ ID NO: 20.  
 DE  
 XX Soybean; plant; RNA-directed RNA polymerase; gene mapping; gene marker;  
 KW plant virus resistance; plant breeding.  
 XX Glycine max.  
 OS  
 XX WO2000060097-A1.  
 PN  
 XX 12-OCT-2000.  
 PD  
 XX 06-APR-2000; 2000WO-US09105.  
 PF  
 XX 07-APR-1999; 99US-0128094.  
 PR  
 XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J, Zhong G;  
 PI  
 XX WPI; 2000-679376/66.  
 DR N-PSDB; AAC63746.  
 XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful  
 PT

PT for controlling gene expression and providing mechanisms to engineer  
 PT plant virus resistance -  
 XX Claim 23; Page 60; 62pp; English.  
 XX The present sequence is part of a plant RNA-directed RNA polymerase  
 CC protein. Polynucleotides encoding RNA-directed RNA polymerases were  
 CC isolated from plant cDNA libraries. They are useful as probes for  
 CC genetically and physically mapping genes, and as markers for  
 CC traits linked to those genes. They are useful for controlling gene  
 CC expression and provide mechanisms to engineer plant virus resistance.  
 CC They are also useful for plant breeding to develop lines with desired  
 CC phenotypes.  
 XX Sequence 60 AA;  
 SQ Query Match 0.8%; Score 9; DB 21; Length 60;  
 Best Local Similarity 100.0%; Pred. No. 0.5;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 891 DFPKGTGVP 899  
 DB 15 DFPKGTGVP 23  
 RESULT 21  
 AAO03533  
 ID AAO03533 standard; Protein; 93 AA.  
 AC  
 AC AAO03533;  
 XX 06-NOV-2001 (first entry)  
 DT  
 XX Human polypeptide SEQ ID NO 17425.  
 DE  
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX Homo sapiens.  
 OS  
 XX WO200164835-A2.  
 PN  
 XX 07-SEP-2001.  
 PD  
 XX 26-FEB-2001; 2001WO-US04927.  
 PF  
 XX 28-FEB-2000; 2000US-0515126.  
 PR 18-MAY-2000; 2000US-0577409.  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Tang YT, Liu C, Drmanac RT;  
 PI  
 XX WPI; 2001-514838/56.  
 DR N-PSDB; AAI83454.  
 DR  
 XX Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -  
 XX Claim 20; SEQ ID NO 17425; 1399pp + Sequence Listing; English.  
 PS  
 XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 93 AA;

QY Query Match 0.7%; Score 8; DB 22; Length 93;  
Db Best Local Similarity 100.0%; Pred. No. 8.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 LKKFFFFEL 134  
Db 51 LKKFFFFEL 58  
|||||

RESULT 22

AAAY70941  
ID AAY70941 standard; Protein; 130 AA.

XX AC AAY70941;

XX XX 09-AUG-2000 (first entry)

XX DE Corn histidine biosynthetic enzyme HisA from clone p0085.csca119r.

XX KW Corn; histidine biosynthetic enzyme; HisA; clone p0085.csca119r;  
KW phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase;  
XX herbicide; fungicide.

XX OS Zea mays.

XX PN WO200024867-A2.

XX PD 04-MAY-2000.

XX PF 21-OCT-1999; 99WO-0524598.

XX PR 23-OCT-1998; 98US-0105409.

XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.

XX PI Abell LM, Allen SM, Cahoon RE, Tao Y;

XX WPI; 2000-350713/30.

XX DR N-PSDB; AAD00287.

XX PT Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide  
PT isomerase, polypeptide (P), useful for discovering compounds which  
PT inhibit isomerase activity and act as herbicides and fungicides -

XX PS Claim 2; Page 38-39; 47pp; English.

XX CC The present sequence is a phosphoribosylformimino-5-  
CC aminoimidazole carboxamide ribotide isomerase (HisA) which is a  
CC histidine biosynthetic enzyme. The sequence is derived from  
CC p0085.csca119r clone isolated from a corn shoot culture cDNA library  
CC p0085. The present sequence is useful in altering the  
CC levels of HisA enzyme in transformed plant cells.

CC It may also be used for discovering compounds that have the ability  
CC to inhibit activity of the isomerase, and therefore act as herbicides  
CC and fungicides.

XX Sequence 130 AA;

QY Query Match 0.7%; Score 8; DB 21; Length 130;  
Db Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 520 TSYVFSFG 527  
Db 99 TSYVFSFG 106  
|||||

RESULT 23

AAU18235

ID AAU18235 standard; Protein; 283 AA.

XX AC AAU18235;

XX XX 21-NOV-2001 (first entry)

XX DE Novel human DNA-binding protein #82.

XX KW Human; DNA-binding protein; histone; chromo domain protein;  
KW chromatin organisation modifier; Y-box binding protein;  
KW DNA organisation; gene transcription; malignant disease;  
KW autoimmune disorder; rheumatic disease; genetic abnormality;  
KW infectious disease; neurological disorder; gene therapy;  
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;  
KW cytostatic.

XX OS Homo sapiens.

XX PN WO200155162-A1.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01305.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 30-JUN-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0216647.

XX PR 11-JUL-2000; 2000US-0216880.

XX PR 11-JUL-2000; 2000US-0217496.

XX PR 14-JUL-2000; 2000US-0218290.

XX PR 26-JUL-2000; 2000US-0220963.

XX PR 26-JUL-2000; 2000US-0220964.

XX PR 14-AUG-2000; 2000US-0224518.

XX PR 14-AUG-2000; 2000US-0224519.

XX PR 14-AUG-2000; 2000US-0225213.

XX PR 14-AUG-2000; 2000US-0225214.

XX PR 14-AUG-2000; 2000US-0225266.

XX PR 14-AUG-2000; 2000US-0225267.

XX PR 14-AUG-2000; 2000US-0225268.

XX PR 14-AUG-2000; 2000US-0225270.

XX PR 14-AUG-2000; 2000US-0225447.

XX PR 14-AUG-2000; 2000US-0225757.

XX PR 14-AUG-2000; 2000US-0225758.

XX PR 14-AUG-2000; 2000US-0225759.

XX PR 18-AUG-2000; 2000US-0226279.

XX PR 22-AUG-2000; 2000US-0226681.

XX PR 22-AUG-2000; 2000US-0226868.

XX PR 22-AUG-2000; 2000US-0227182.

XX PR 23-AUG-2000; 2000US-0227009.

XX PR 30-AUG-2000; 2000US-0228924.

XX PR 01-SEP-2000; 2000US-0229287.

XX PR 01-SEP-2000; 2000US-0229343.

XX PR 01-SEP-2000; 2000US-0229344.

XX PR 01-SEP-2000; 2000US-0229345.

XX PR 05-SEP-2000; 2000US-0229509.

XX PR 05-SEP-2000; 2000US-0229513.

XX PR 06-SEP-2000; 2000US-0230437.

XX PR 06-SEP-2000; 2000US-0230438.

XX PR 08-SEP-2000; 2000US-0231242.



KW Wheat; histidine biosynthetic enzyme; HisA; clone wleln.pk0018.b10;  
KW phosphoribosylformimino-5-aminimidazole carboxamide ribotide isomerase;  
KW herbicide; fungicide.

XX Triticum aestivum.  
OS WO200024867-A2.  
PN  
XX  
PD 04-MAY-2000.  
XX

PF 21-OCT-1999; 99WO-US24698.  
XX  
XX 23-OCT-1998; 98US-0105409.  
PR  
XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
PA

XX Abell LM, Allen SM, Cahoon RE, Tao Y;  
PI  
XX WPI; 2000-350713/30.  
DR N-PSDB; AAD00293.  
DR

XX Phosphoribosylformimino-5-aminimidazole carboxamide ribotide  
PT isomerase, polypeptide (P), useful for discovering compounds which  
PT inhibit isomerase activity and act as herbicides and fungicides -  
XX  
XX Claim 2; Page 46-47; 47pp; English.

XX The present sequence is a phosphoribosylformimino-5-  
CC aminoimidazole carboxamide ribotide isomerase (HisA) which is a  
CC histidine biosynthetic enzyme. The sequence is derived from  
CC wleln.pk0018.b10 clone isolated from a wheat leaf cDNA library wleln.  
CC The present sequence is useful in altering the  
CC levels of HisA enzyme in transformed plant cells.  
CC It may also be used for discovering compounds that have the ability  
CC to inhibit activity of the isomerase, and therefore act as herbicides  
CC and fungicides.

XX SQ Sequence 327 AA;  
Query Match 0.7%; Score 8; DB 21; Length 327;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 520 TSVVFSFG 527  
DB 172 TSVVFSFG 179  
|||||

RESULT 25  
AAV09307  
ID AAY09307 standard; Protein; 341 AA.  
XX  
AC AAY09307;

XX\* 09-JUL-1999 (first entry)  
PI  
XX Human Vpr interacting protein.  
DE

XX Human; Vpr interacting protein; VIP; hVIP; cellular receptor; HIV-1;  
KW G2/M phase transition; cell cycle; cancer; hyperproliferative disease;  
KW autoimmune disease.

XX Homo sapiens.  
OS WO9919359-A1.  
PN  
XX 22-APR-1999.  
PD  
XX 09-OCT-1998; 98WO-US21432.  
XX  
XX 10-OCT-1997; 97US-0949202.  
PR  
XX (UYPE-) UNIV PENNSYLVANIA.

XX Ayyavoo V, Mahalingam S, Patel M, Weiner DB;  
PI  
XX WPI; 1999-277596/23.  
DR N-PSDB; AAX55871.  
XX  
XX Human Vpr interacting protein for, e.g. treatment of cancer  
XX  
XX Claim 1; Page 51-52; 58pp; English.

XX The present sequence represents human Vpr interacting protein (hVIP).  
CC DNA encoding hVIP can be used to transform host cells to express the  
CC protein. Sequences complementary to the DNA can be used in methods for  
CC inhibiting expression of hVIP, treating an individual who has cancer,  
CC and for inhibiting hVIP activity in a cell. The hVIP can also be used  
CC to treat hyperproliferative diseases and some autoimmune diseases.

XX SQ Sequence 341 AA;  
Query Match 0.7%; Score 8; DB 20; Length 341;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 FFFFLSSG 137  
DB 26 FFFFLSSG 33  
|||||

RESULT 26  
AAV09308  
ID AAY09308 standard; Protein; 341 AA.  
XX  
AC AAY09308;

XX 09-JUL-1999 (first entry)  
DT  
XX Alternative human Vpr interacting protein.  
DE

XX Human; Vpr interacting protein; VIP; hVIP; cellular receptor; HIV-1;  
KW G2/M phase transition; cell cycle; cancer; hyperproliferative disease;  
KW autoimmune disease.

XX Homo sapiens.  
OS WO9919359-A1.  
PN  
XX 22-APR-1999.  
PD  
XX 09-OCT-1998; 98WO-US21432.  
PF  
XX 10-OCT-1997; 97US-0949202.  
PR  
XX (UYPE-) UNIV PENNSYLVANIA.

XX Ayyavoo V, Mahalingam S, Patel M, Weiner DB;  
PI  
XX WPI; 1999-277596/23.  
DR N-PSDB; AAX55872.

XX Human Vpr interacting protein for, e.g. treatment of cancer  
XX  
XX Example 1; Page 54-56; 58pp; English.

XX The present sequence represents human Vpr interacting protein (hVIP).  
CC DNA encoding hVIP can be used to transform host cells to express the  
CC protein. Sequences complementary to the DNA can be used in methods for  
CC inhibiting expression of hVIP, treating an individual who has cancer,  
CC and for inhibiting hVIP activity in a cell. The hVIP can also be used  
CC to treat hyperproliferative diseases and some autoimmune diseases.

XX SQ Sequence 341 AA;  
Query Match 0.7%; Score 8; DB 20; Length 341;





PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249399.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 11-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI  
 XX  
 XX WPI; 2001-465557/50.  
 DR N-PSDB; AAS29031.  
 DR  
 XX  
 XX Nucleic acid molecules encoding human secreted chromosomal binding  
 FT proteins, used in preventing, treating or ameliorating a disorder, e.g.  
 FT Alzheimer's and Parkinson's diseases and cancers -  
 PT  
 XX  
 XX Claim 11; SEQ ID No 140; 561pp; English.  
 PS  
 XX  
 XX The present invention relates to the isolation of novel DNA-binding  
 CC proteins, and cDNA (AAS29030-AAS29157) and genomic sequences encoding  
 CC for these proteins. DNA-binding proteins such as histones, chromo  
 CC (chromatin organisation modifier) domain proteins, and Y-box binding  
 CC proteins may contribute to diseases resulting from aberrant DNA  
 CC organisation and/or gene transcription. The sequences of the invention  
 CC are useful in screening assays to identify antagonists and/or agonists  
 CC that may enhance or block activities mediated by DNA-binding proteins.  
 CC Blockers of DNA-binding proteins may be useful in treating disorders  
 CC such as malignant diseases (e.g. cancer), autoimmune disorders  
 CC (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid  
 CC arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious  
 CC diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's  
 CC disease). The polynucleotide sequences of the invention may also be  
 CC used in gene therapy. AAU18154-AAU18281 represent novel DNA-binding  
 CC proteins.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 477 AA;  
 Query Match 0.7%; Score 8; DB 22; Length 477;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 739 SRSTNSNF 746  
 DB 114 SRSTNSNF 121  
 RESULT 28  
 AAO21854  
 ID AAO21854 standard; Protein; 641 AA.  
 XX

AC AAO21854;  
 XX  
 DT 13-SEP-2002 (first entry)  
 XX  
 DE Isoprenoid related protein sequence SEQ ID No 25.  
 XX  
 KW Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DDS;  
 KW decaprenyl diphosphate synthase.  
 XX  
 OS Rhodobacter capsulatus.  
 XX  
 PN WC200226933-A2.  
 XX  
 PD 04-APR-2002.  
 XX  
 PF 28-SEP-2001; 2001WO-US30328.  
 XX  
 PR 29-SEP-2000; 2000US-236580P.  
 XX  
 PA (CRGI ) CARGILL INC.  
 XX  
 PI Gokarn R, Jessen H, Zidwick MJ;  
 XX  
 DR WPI; 2002-416480/44.  
 XX  
 PT Substantially pure polypeptides having e.g.,  
 PT 1-deoxyxylulose-5-phosphate synthase activity, useful for the  
 PT production of isoprenoids, especially CoQ(10) -  
 XX  
 PS Disclosure; Fig 6; 246pp; English.  
 XX  
 CC The invention relates to methods and materials for the production of  
 CC isoprenoids. More particularly the invention provides isolated nucleic  
 CC acids, substantially pure polypeptides, host cells, and methods for  
 CC producing various isoprenoid compounds. The polypeptides are useful for  
 CC the production of isoprenoids, especially CoQ(10). Expressing the pure  
 CC polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (DXS)  
 CC activity or decaprenyl diphosphate synthase (DDS) activity, is useful for  
 CC increasing production of CoQ(10) in a cell having endogenous DDS  
 CC activity. This sequence represents a protein relating to the isoprenoid  
 CC production of the invention.  
 XX  
 SQ Sequence 641 AA;  
 Query Match 0.7%; Score 8; DB 23; Length 641;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 465 GTANDIR 472  
 DB 615 GTANDIR 622  
 RESULT 29  
 ABB93071  
 ID ABB93071 standard; Protein; 830 AA.  
 XX  
 AC ABB93071;  
 XX  
 DT 31-MAY-2002 (first entry)  
 XX  
 DE Herbicidally active polypeptide SEQ ID NO 2282.  
 DE  
 XX Herbicidal; plant; agriculture; herbicide.  
 KW  
 XX Arabidopsis thaliana.  
 OS  
 XX WO200210210-A2.  
 PN  
 XX  
 PD 07-FEB-2002.  
 XX  
 PF 28-AUG-2001; 2001WO-EP09892.  
 XX



CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1600 AA;

Query Match 0.7%; Score 8; DB 22; Length 1600;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 68 RLYFGSSY 75  
 |||||  
 DB 434 RLYFGSSY 441

RESULT 32

ABG12275  
 ID ABG12275 standard; Protein; 1600 AA.

XX AC ABG12275;

DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #12266.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI: 2001-639362/73.

XX DR N-PSDB; AAS76462.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -

XX PS Claim 20; SEQ ID No 42634; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and

XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1600 AA;

Query Match 0.7%; Score 8; DB 22; Length 1600;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 68 RLYFGSSY 75  
 |||||  
 DB 434 RLYFGSSY 441

RESULT 33

ABG20004  
 ID ABG20004 standard; Protein; 1600 AA.

XX AC ABG20004;

DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #19995.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI: 2001-639362/73.

XX DR N-PSDB; AAS84191.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -

XX PS Claim 20; SEQ ID No 50363; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and

XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 XX CC and gene mapping, and in recombinant production of (II). The

XX CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1600 AA;

Query Match 0.7%; Score 8; DB 22; Length 1600;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 RLYFGSSY 75  
 |||||  
 Db 434 RLYFGSSY 441

# RESULT 34

ABG09420  
 ID ABG09420 standard; Protein; 2168 AA.

XX AC ABG09420;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #9411.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS73607.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -

XX PS Claim 20; SEQ ID NO 39779; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2168 AA;

Query Match 0.7%; Score 8; DB 22; Length 2168;

Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 RLYFGSSY 75  
 |||||  
 Db 949 RLYFGSSY 956

# RESULT 35

ABG09727  
 ID ABG09727 standard; Protein; 2168 AA.

XX AC ABG09727;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #9718.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS73914.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -

XX PS Claim 20; SEQ ID NO 40086; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2168 AA;  
 Query Match 0.7%; Score 8; DB 22; Length 2168;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 68 RLYFGSSY 75  
 Db 949 RLYFGSSY 956  
 |||||

RESULT 36  
 ABG12287  
 ID ABG12287 standard; Protein; 2176 AA.  
 XX AC ABG12287;  
 XX DT 18-FEB-2002 (first entry)  
 XX DE Novel human diagnostic protein #12278.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.  
 XX PN WO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US08631.  
 XX PR 31-MAR-2000; 2000US-0540217.  
 XX PR 23-AUG-2000; 2000US-0649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Drmanac RT, Liu C, Tang YT;  
 XX DR WPI; 2001-639362/73.  
 XX DR N-PSDB; AAS76474.  
 XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX PS Claim 20; SEQ ID NO 42646; 103pp; English.  
 XX CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2176 AA;  
 Query Match 0.7%; Score 8; DB 22; Length 2176;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 68 RLYFGSSY 75  
 Db 949 RLYFGSSY 956  
 |||||

RESULT 37  
 ABG20022  
 ID ABG20022 standard; Protein; 2176 AA.  
 XX AC ABG20022;  
 XX DT 13-FEB-2002 (first entry)  
 XX DE Novel human diagnostic protein #20013.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.  
 XX PN WO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US08631.  
 XX PR 31-MAR-2000; 2000US-0540217.  
 XX PR 23-AUG-2000; 2000US-0649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Drmanac RT, Liu C, Tang YT;  
 XX DR WPI; 2001-639362/73.  
 XX DR N-PSDB; AAS84209.  
 XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX PS Claim 20; SEQ ID NO 50381; 103pp; English.  
 XX CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG0377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 2176 AA;

Query Match 0.7%; Score 8; DB 22; Length 2176;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 68 RLYFGSSV 75  
 |||||  
 Db 949 RLYFGSSV 956

RESULT 38  
 AAU10011  
 ID AAU10011 standard; Peptide; 18 AA.  
 XX  
 AC AAU10011;  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE Tomato RNA-directed RNA polymerase (RdRP) 130 kDa-15 AS sequence.  
 XX  
 KW Tomato; RdRP: RNA-directed RNA polymerase; in vitro transcription;  
 KW cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;  
 KW transgenic plant; transgenic animal; cancer; viral infection;  
 KW immunoprecipitation; immunolocalisation; 130 kDa-15; gene therapy.  
 XX  
 OS Lycopodium esculentum.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 15  
 FT /label= Asn, Lys  
 PN US2001023067-A1.  
 XX  
 PD 20-SEP-2001.  
 XX  
 PF 08-FEB-2001; 2001US-0782874.  
 XX  
 PR 05-MAR-1997; 97US-0811583.  
 XX  
 PA (WASS/) WASSENEGGER M.  
 PA (RIED/) RIEDEL L.  
 XX  
 PI Wassenegger M, Riedel L, Schleibel W, Sanger HL;  
 XX  
 XX WPI; 2001-595798/67.  
 DR N-PSDB; AAS17849.  
 XX  
 PT New nucleic acid molecule encoding a polypeptide having the enzymatic  
 PT activity of RNA-directed RNA polymerase, for modulating gene expression  
 PT and treating cancer and virus infection in human and animals -  
 XX  
 PS Example 3; Fig 6; 34pp; English.  
 XX  
 CC This sequence represents a tomato RNA-directed RNA polymerase (RdRP)  
 CC peptide sequence used to create a degenerate nucleotide sequence and  
 CC degenerate PCR primers used to amplify the RdRP cDNA sequence of the  
 CC invention. The invention comprises the nucleic acid and protein  
 CC sequences of RdRP, the protein of the invention can catalyze in vitro  
 CC transcription of short single stranded RNAs into DNA molecules,  
 CC this transcription can be either primed by RNA or DNA oligonucleotides  
 CC or be unprimed. The protein may have cytosolic or virucide activities.  
 CC The sequences of the invention may be used in gene therapy or as an RNA  
 CC directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template

CC nucleic acid molecule derived from a nucleic acid molecule which causes  
 CC a disease are useful for treating a disease caused by the undesired  
 CC expression or overexpression of a nucleic acid molecule in a human, rat  
 CC or mouse, by administering the molecules. This system can be used in the  
 CC preparation of a pharmaceutical composition and for inhibiting  
 CC expression of any desired gene by transferring the RdRP system to  
 CC organisms that either lack a comparable mechanism or do not sufficiently  
 CC express their own RdRP. An antibody or an antagonist or inhibitor to the  
 CC protein are useful for inhibiting RNA directed RNA synthesis and for  
 CC ensuring stable heterologous, gene expression in transgenic organisms.  
 CC The sequence is useful for probes and/or for the control of gene  
 CC expression, as primers for amplification of nucleic acid molecules and  
 CC as tools for the detection of expression of the cDNA molecules.  
 CC Additionally, nucleotide and protein sequences are useful for  
 CC suppression of undesired gene expression in humans and animals. The RdRP  
 CC is useful as a therapeutic agent for the control of cancer and virus  
 CC infection in humans and animals and the antibody is useful for  
 CC immunoprecipitation or immunolocalisation of the protein, identification  
 CC of polypeptides interacting with it and screening expression libraries.  
 XX  
 SQ Sequence 18 AA;

Query Match 0.6%; Score 7; DB 22; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 640 KAQEALE 646  
 |||||  
 Db 1 KAQEALE 7

RESULT 39  
 ABB43113  
 ID ABB43113 standard; Peptide; 18 AA.  
 XX  
 AC ABB43113;  
 XX  
 DT 04-FEB-2002 (first entry)  
 XX  
 DE Peptide #10619 encoded by human foetal liver single exon probe.  
 XX  
 KW Human; foetal liver; gene expression; single exon nucleic acid probe.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157277-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PR 30-JAN-2001; 2001WO-US00669.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 XX WPI; 2001-483447/52.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human fetal liver -  
 XX  
 PS Claim 27; SEQ ID NO 35749; 639pp + sequence listing; English.  
 XX  
 CC The invention relates to a single exon nucleic acid probe for  
 CC measuring human gene expression in a sample derived from human foetal  
 CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human  
CC fetal liver. The present sequence is a peptide encoded by a single exon  
CC nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX SQ Sequence 18 AA;

Query Match 0.6%; Score 7; DB 22; Length 18;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 572 KLSLRKS 578  
DB 12 KLSLRKS 18  
|||||||

RESULT 40  
AAM64015  
ID AAM64015 standard; Protein; 18 AA.

XX  
AC AAM64015;

XX DT 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe encoded protein SEQ ID NO: 36120.

XX Human: brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer.

XX OS Homo sapiens.

XX PN WO200157275-A2.

XX PD 09-AUG-2001.

XX PE 30-JAN-2001; 2001WO-US00667.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX XX WPI; 2001-483445/52.

XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains -  
XX Example 4; SEQ ID NO: 36120; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX brain. They can be used to measure gene expression in brain cell samples,  
XX which may enable the diagnosis and improved treatment of nervous system  
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
XX epilepsy and cancers. The present sequence is a protein encoded by one of  
XX the probes of the invention.

XX SQ Sequence 18 AA;

Query Match 0.6%; Score 7; DB 22; Length 18;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 572 KLSLRKS 578  
DB 12 KLSLRKS 18  
|||||||

RESULT 41  
AAM76838  
ID AAM76838 standard; Protein; 18 AA.

XX  
AC AAM76838;

XX DT 06-NOV-2001 (first entry)

XX Human bone marrow expressed probe encoded protein SEQ ID NO: 37144.  
XX Human: bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma.

XX OS Homo sapiens.

XX PN WO200157276-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00668.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human bone marrow -  
XX Example 4; SEQ ID NO: 37144; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX bone marrow. They can be used to measure gene expression in bone marrow  
XX samples, which may enable the improved diagnosis and treatment of cancers  
XX such as lymphoma, leukaemia and myeloma. The present sequence is a  
XX protein encoded by one of the probes of the invention.

XX SQ Sequence 18 AA;

Query Match 0.6%; Score 7; DB 22; Length 18;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 572 KLSLRKS 578  
DB 12 KLSLRKS 18  
|||||||

RESULT 42  
AAM36944  
ID AAM36944 standard; Protein; 18 AA.

XX  
AC AAM36944;

XX DT 17-OCT-2001 (first entry)

XX Peptide #10981 encoded by probe for measuring placental gene expression.

XX DE



KW Probe; microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder.

XX Homo sapiens.

PN WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

XX Claim 27; SEQ ID No 37213; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP;

XX see AA131315-AA157546). The present sequence is a peptide encoded by one

XX such probe. The probes are useful for producing a microarray for

XX predicting, measuring and displaying gene expression in samples derived

XX from human placenta. The probes are useful for antenatal diagnosis of

XX human genetic disorders.

XX Sequence 18 AA;

SQ Query Match

Best Local Similarity 0.6%; Score 7; DB 22; Length 18;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 572 KLSLRKS 578

DB 12 KLSLRKS 18

RESULT 43

AAE00904

ID AAE00904 standard; peptide; 18 AA.

XX AAE00904;

XX 04-JUL-2001 (first entry)

XX Peptide #2 obtained by micro-sequencing Rdrp protein.

XX Tomato; gene therapy; RNA-directed RNA polymerase; Rdrp; gene expression;

XX transgenic plant; tissue culture; plant breeding; therapy; C-protein.

XX Lycopersicon esculentum.

XX Key Location/Qualifiers

FT Misc-difference 6 /note= "Encoded by YTN"

XX US6218142-B1.

XX 17-APR-2001.

XX 05-MAR-1997; 97US-0811583.

XX 05-MAR-1997; 97US-0811583.

XX (WASS/) WASENEGGER M.

XX (RIED/) RIEDEL L.

XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;

XX WPI; 2001-289830/30.

XX N-PSDB; AAD04382.

XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA

XX polymerase enzymatic activity, useful in modulating gene expression in

XX plants, humans and animals, as well as in plant cell/tissue cultures or

XX plant breeding -

XX Example 3; Fig 6; 31pp; English.

XX The present sequence is a peptide encoded by an oligonucleotide which is

XX used to design RNA-directed RNA polymerase (Rdrp)-specific PCR primers.

XX These PCR primers are used to amplify a cDNA encoding tomato C-protein

XX having Rdrp activity. The peptide is obtained by micro-sequencing Rdrp

XX protein. C-protein is capable of RNA-directed RNA synthesis, thus using

XX RNA as a template for synthesizing complementary RNA molecules. Rdrp

XX nucleic acid is useful for modulating gene expression in plants, humans

XX and animals. This may lead to various physiological, developmental and/or

XX morphological changes. Transgenic plants containing Rdrp nucleic acid is

XX especially useful in plant cell or tissue cultures and in plant breeding.

XX Rdrp is useful in gene therapy, particularly for treating a disease that

XX is caused by the undesirable expression or overexpression of a gene.

XX Sequence 18 AA;

SQ Query Match

Best Local Similarity 0.6%; Score 7; DB 22; Length 18;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 640 KAQEALE 646

DB 1 KAQEALE 7

RESULT 44

ABB4400

ID ABB4400 standard; Peptide; 24 AA.

XX ABB4400;

XX 04-FEB-2002 (first entry)

XX Peptide #11906 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX Homo sapiens.

XX WC200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.



```
XX SQ Sequence 24 AA;
Query Match 0.6%; Score 7; DB 22; Length 24;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 209 LSSSLCL 215
Db 7 LSSSLCL 13

RESULT 47
AAY18363
ID AAY18363 standard; peptide; 26 AA.
XX AC AAY18363;
XX DT 20-AUG-1999 (first entry)
XX DE Human chromogranins A fragment.
XX KW Chromogranins A; CgA; human; antibody production; detection.
XX OS Homo sapiens.
XX PN JP11153598-A.
XX PD 08-JUN-1999.
XX PF 20-NOV-1997; 97JP-0336452.
XX PR 20-NOV-1997; 97JP-0336452.
XX PA (YANA-) YANAIHARA KENYUSHO KK.
XX WPI; 1999-390410/33.
New antibody - has reactivity specific to human chromogranins A
(CgA)
Claim 1; Page 2; 10pp; Japanese.
This sequence represents a fragment of human chromogranins A (CgA). The
invention relates to an antibody having a reactivity specific to human
CgA, which is prepared by immunising a warm-blooded animal, other than
human, using this peptide. The antibody is useful in a method for
detecting and/or measuring human CgA. The method can measure and detect
human CgA with a high sensitivity.
XX SQ Sequence 31 AA;
Query Match 0.6%; Score 7; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 568 DSSMKLS 574
Db 8 DSSMKLS 14
|||||
RESULT 49
AAB33978
ID AAB33978 standard; Peptide; 36 AA.
XX AC AAB33978;
XX DT 02-FEB-2001 (first entry)
XX DE Human secreted protein encoded by cDNA #17.
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX OS Homo sapiens.
XX PN WO2000056765-A1.
XX PD 28-SEP-2000.
XX PF 16-MAR-2000; 2000WO-US06923.
XX PR 19-MAR-1999; 99US-0125364.
XX PR 08-DEC-1999; 99US-0169623.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM, Komatsoulis G;
XX WPI; 2000-502215/57.
XX N-PSDB; AAC59408.
Nucleic acid molecules encoding human secreted proteins, used in
preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
```

PT Parkinson's diseases and cancers -  
 XX Claim 11; Page 363; 410pp; English.  
 XX Sequences AAB33963-B34006 represent the amino acid sequences of 48  
 CC human secreted proteins encoded by the genes AAC59322-C59439. The genes  
 CC and proteins are useful for preventing, ameliorating or treating  
 CC medical conditions, e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
 CC ovarian cancer, and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and  
 CC ulcerative colitis; (c) cardiovascular disorders such as myocardial  
 CC ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral  
 CC anoxia and epilepsy; and (f) infectious diseases such as viral,  
 CC bacterial, fungal and parasitic infections.  
 XX Sequence 36 AA;  
 SQ  
 Query Match 0.6%; Score 7; DB 21; Length 36;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 130 FFFFLSS 136  
 |||||  
 Db 20 FFFFLSS 26

RESULT 50  
 AAB25008  
 ID AAB25008 standard; Peptide; 40 AA.  
 XX  
 AC AAB25008;  
 XX  
 DT 27-NOV-2000 (first entry)  
 XX  
 DE Plant SDF encoded polypeptide sequence SEQ List 2 NO:38.  
 XX  
 KW Plant; corn; Arabidopsis thaliana; sequence-determined DNA fragment;  
 KW SDF; genetic mapping; identification; promoter; structural gene; UTR;  
 KW untranslated region; expression control.  
 XX  
 OS Plant.  
 XX  
 PN WO2000040695-A2.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 07-JAN-2000; 2000WO-US00466.  
 XX  
 PR 08-JAN-1999; 99US-0115293.  
 XX  
 PA (CERE-) CERES INC.  
 XX  
 PI Alexandrov N, Brover V, Chen X, Subramanian G, Troukhan ME;  
 PI Zheng L;  
 XX  
 DR WPI; 2000-465970/40.  
 XX  
 PT New corn plant and Arabidopsis thaliana sequence-determined DNA  
 PT fragments, useful for expressing gene products and for controlling  
 PT expression of a target gene -  
 XX  
 PS Claim 14; Page 615; 673pp; English.  
 XX  
 CC The present invention describes polynucleotides, such as complete cDNA  
 CC sequences and/or sequences of genomic DNA encompassing complete genes,  
 CC portions of genes, and/or intergenic regions, collectively referred to  
 CC as sequence-determined DNA fragments (SDFs), from corn plants and

CC Arabidopsis thaliana. The SDFs are promoters, structural genes,  
 CC untranslated regions (UTRs), or 3' termination sequences. They can be  
 CC used for expressing a gene product and controlling expression of a  
 CC target gene, either as a promoter, a structural gene, an UTR or as a  
 CC 3' termination sequence. They are also useful as tools for genetic  
 CC mapping, and identification of a particular individual plant or for  
 CC clustering a group of plants with a common trait. AAB78433 to AAB78630  
 CC and AAB24605 to AAB25099 represent the specifically claimed  
 CC polynucleotide sequences and polypeptides encoded by them given in the  
 CC present invention.  
 XX  
 SQ Sequence 40 AA;  
 Query Match 0.6%; Score 7; DB 21; Length 40;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 130 FFFFLSS 136  
 |||||  
 Db 19 FFFFLSS 25

Search completed: November 6, 2002, 03:33:28  
 Job time : 164 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 03:36:48 ; Search time 83 Seconds  
(without alignments)  
394.905 Million cell updates/sec

Title: US-09-782-874-2  
Perfect score: 1114  
Sequence: 1 MCKTIQVFGFPYLLSAEVVK.....RPVLNLSLRQAQLSHRLVLIK 1114

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database : Issued Patents\_AA.\*  
1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/prodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1114	100.0	1114	4	US-08-811-583-2
2	218	19.6	218	4	US-08-811-583-3
3	17	1.5	17	4	US-08-811-583-12
4	14	1.3	14	4	US-08-811-583-10
5	14	1.3	14	4	US-08-811-583-11
6	10	0.9	13	4	US-08-811-583-13
7	8	0.7	341	3	US-08-949-202-2
8	8	0.7	341	3	US-08-949-202-4
9	8	0.7	341	4	US-09-418-175-2
10	8	0.7	341	4	US-09-418-175-4
11	8	0.7	341	4	US-08-529-245-2
12	8	0.7	341	4	US-09-529-245-4
13	8	0.7	506	4	US-09-457-040B-17
14	7	0.6	149	4	US-09-134-001C-3012
15	7	0.6	154	4	US-09-134-001C-3132
16	7	0.6	165	4	US-09-455-960-7
17	7	0.6	220	4	US-08-297-431B-31
18	7	0.6	221	4	US-08-297-431B-2
19	7	0.6	221	4	US-08-297-431B-4
20	7	0.6	221	4	US-08-297-431B-6
21	7	0.6	221	4	US-08-297-431B-8
22	7	0.6	221	4	US-08-297-431B-10
23	7	0.6	221	4	US-08-297-431B-12
24	7	0.6	221	4	US-08-297-431B-14
25	7	0.6	221	4	US-08-297-431B-16
26	7	0.6	221	4	US-08-297-431B-18
27	7	0.6	221	4	US-08-297-431B-20

28	7	0.6	320	1	US-08-565-386-17	Sequence 17, Appl
29	7	0.6	322	2	US-08-576-626A-33	Sequence 33, Appl
30	7	0.6	394	4	US-08-791-115B-27	Sequence 27, Appl
31	7	0.6	403	4	US-08-791-115B-1	Sequence 1, Appl
32	7	0.6	403	4	US-08-791-115B-23	Sequence 23, Appl
33	7	0.6	403	4	US-08-791-115B-25	Sequence 25, Appl
34	7	0.6	430	4	US-08-791-115B-7	Sequence 7, Appl
35	7	0.6	470	2	US-08-933-821-6	Sequence 6, Appl
36	7	0.6	470	3	US-08-960-507-6	Sequence 6, Appl
37	7	0.6	470	4	US-09-136-828-6	Sequence 6, Appl
38	7	0.6	470	4	US-09-332-928A-6	Sequence 6, Appl
39	7	0.6	470	4	US-09-136-801-6	Sequence 6, Appl
40	7	0.6	470	4	US-09-332-929-6	Sequence 6, Appl
41	7	0.6	501	3	US-08-968-563-15	Sequence 15, Appl
42	7	0.6	501	4	US-08-969-683A-15	Sequence 15, Appl
43	7	0.6	501	4	US-09-297-928-11	Sequence 11, Appl
44	7	0.6	565	4	US-08-906-156A-12	Sequence 12, Appl
45	7	0.6	577	4	US-09-183-266A-49	Sequence 49, Appl
46	7	0.6	645	4	US-08-791-115B-6	Sequence 6, Appl
47	7	0.6	742	4	US-08-791-115B-5	Sequence 5, Appl
48	7	0.6	834	4	US-09-143-571-29	Sequence 29, Appl
49	7	0.6	884	4	US-09-741-150-4	Sequence 4, Appl
50	7	0.6	971	2	US-08-724-354D-22	Sequence 22, Appl

## ALIGNMENTS

RESULT 1  
US-08-811-583-2  
; Sequence 2, Application US/08811583  
; Patent No. 6218142  
; GENERAL INFORMATION:  
; APPLICANT: Wassenecker, Michael  
; APPLICANT: Riedel, Leonhard  
; APPLICANT: Schiebel, Winfried  
; APPLICANT: Sanger, Heinz  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN  
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RDRP)  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & NEAVE  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/811.583  
; FILING DATE: 05-MAR-1997  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley, James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: MPG-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1114 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-811-583-2

Query Match

100.0%; Score 1114; DB 4; Length 1114;

Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKTIQVGFPPYLLSABVWKSFLKTYGTGTCALVQKSGSRAFAKQVQFADNISADK 60  
DB 1 MGKTIQVGFPPYLLSABVWKSFLKTYGTGTCALVQKSGSRAFAKQVQFADNISADK 60

QY 61 IITLANNRLYFGSSYLKAWEMKTDIVOLRAYVQDMGDTILNFCQISDDKFAVLGSTEVS 120  
DB 61 IITLANNRLYFGSSYLKAWEMKTDIVOLRAYVQDMGDTILNFCQISDDKFAVLGSTEVS 120

QY 121 IQFGIGLKKFFFLSSGADYKQLQSVENIOWVVLHRRPGYQNAQFLLIOLFGAPRIYKRL 180  
DB 121 IQFGIGLKKFFFLSSGADYKQLQSVENIOWVVLHRRPGYQNAQFLLIOLFGAPRIYKRL 180

QY 181 ENSCYSPFKETPDQWVRTTDFPSSWIGLSSSLCLOFRGRVRLPNFEESFFHYAERENNI 240  
DB 181 ENSCYSPFKETPDQWVRTTDFPSSWIGLSSSLCLOFRGRVRLPNFEESFFHYAERENNI 240

QY 241 TLOGTFFVYSOKSALPNVQPPGEGISPIKILFKISLVOHGCIPGAPALNVFFRLVDP 300  
DB 241 TLOGTFFVYSOKSALPNVQPPGEGISPIKILFKISLVOHGCIPGAPALNVFFRLVDP 300

QY 301 RRRNVACIEHALEKLYIKKCCYDPVRLTEQYDGYLKGQPKPSITLDDGLVYVRRV 360  
DB 301 RRRNVACIEHALEKLYIKKCCYDPVRLTEQYDGYLKGQPKPSITLDDGLVYVRRV 360

QY 361 LVTPCKYVFCGPEVNSNRVLRNYSIEDINFLRVSFVDEWEKLYSTDLLPKASTGSGYR 420  
DB 361 LVTPCKYVFCGPEVNSNRVLRNYSIEDINFLRVSFVDEWEKLYSTDLLPKASTGSGYR 420

QY 421 TNYIERILSLRKGFVIGDKKFFFLAFSSQLRDNVSMFASRPLGTANDIRAWMGDFSQ 480  
DB 421 TNYIERILSLRKGFVIGDKKFFFLAFSSQLRDNVSMFASRPLGTANDIRAWMGDFSQ 480

QY 481 IKNVAKYARLQSGFSGSRSTLSVLREIEVDPVKVHGTSYVFSDGIGKISGDFAHRA 540  
DB 481 IKNVAKYARLQSGFSGSRSTLSVLREIEVDPVKVHGTSYVFSDGIGKISGDFAHRA 540

QY 541 SKCGLQTPSAFQIRYGGYGVGVDPDSSMKLSLRKMSKYESDNIKLDVLGWSKYQPC 600  
DB 541 SKCGLQTPSAFQIRYGGYGVGVDPDSSMKLSLRKMSKYESDNIKLDVLGWSKYQPC 600

QY 601 YLNQLITLSTLVQKVEVLEQKQKAVDQDAILHDSLKAQALELMSPGENTNLIKAM 660  
DB 601 YLNQLITLSTLVQKVEVLEQKQKAVDQDAILHDSLKAQALELMSPGENTNLIKAM 660

QY 661 LNCGYKPDAPFLSMQLQTRASKLLDLRTSRIFIPNGRTMGCLEDSRTLEYGOVFO 720  
DB 661 LNCGYKPDAPFLSMQLQTRASKLLDLRTSRIFIPNGRTMGCLEDSRTLEYGOVFO 720

QY 721 FTGAGHGEFSDDLHPFNNSRSTNSNFIKGNVYVAKNPLHGDIVRLKAVNVRALHHV 780  
DB 721 FTGAGHGEFSDDLHPFNNSRSTNSNFIKGNVYVAKNPLHGDIVRLKAVNVRALHHV 780

QY 781 DCVVPQKGRPHNECSGSDLDGDIYFVCWQDMIPROVQPMEXPPAPSIQLDHDTV 840  
DB 781 DCVVPQKGRPHNECSGSDLDGDIYFVCWQDMIPROVQPMEXPPAPSIQLDHDTV 840

QY 841 EEVEEYFTNYIVNDSLGIIANAHVVFADREPDMAMSDPCCKKLAELFSIAYDFPKTGPAP 900  
DB 841 EEVEEYFTNYIVNDSLGIIANAHVVFADREPDMAMSDPCCKKLAELFSIAYDFPKTGPAP 900

QY 901 IPSQLRPKEYPDFMDKPKTSYISERVIGKLFVKVKQAPQASSIAFTTRDVARRSYDAD 960  
DB 901 IPSQLRPKEYPDFMDKPKTSYISERVIGKLFVKVKQAPQASSIAFTTRDVARRSYDAD 960

QY 961 MEVDGFEDYIDEAFDYKTEYDNKLGNDYIGIKTEAELLSSGIMKASKTFDRRKDAEAI 1020  
DB 961 MEVDGFEDYIDEAFDYKTEYDNKLGNDYIGIKTEAELLSSGIMKASKTFDRRKDAEAI 1020

QY 1021 SVAVRALKEARAFKRRNDIDMLPKASAWHYHTYHPTWGCYNQGLKRAHFISFPWCV 1080  
DB 1021 SVAVRALKEARAFKRRNDIDMLPKASAWHYHTYHPTWGCYNQGLKRAHFISFPWCV 1080

DB 1021 SVAVRALKEARAFKRRNDIDMLPKASAWHYHTYHPTWGCYNQGLKRAHFISFPWCV 1080

QY 1081 YDQLIQIKKDKARNRPVLNLSLRAQLSHRLVLK 1114  
DB 1081 YDQLIQIKKDKARNRPVLNLSLRAQLSHRLVLK 1114

RESULT 2  
US-08-811-583-3  
; Sequence 3, Application US/08811583  
; Patent No. 6218142  
; GENERAL INFORMATION:  
; APPLICANT: Wassenegger, Michael  
; APPLICANT: Riedel, Leonhard  
; APPLICANT: Schiebel, Winfried  
; APPLICANT: Sanger, Helmut  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMIC ACTIVITY OF AN  
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RRP)  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & NEAVE  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/811,583  
; FILING DATE: 05-MAR-1997  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley, James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: MPG-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 218 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: Protein  
; US-08-811-583-3

Query Match 19.6%; Score 218; DB 4; Length 218;  
Best Local Similarity 100.0%; Pred. No. 7.4e-211;  
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 RTMMGLCDLSRTLEYGOVFOQTGAGHGEFSDDLHPFNNSRSTNSNFIKGNVYVAKNPC 759  
DB 1 RTMMGLCDLSRTLEYGOVFOQTGAGHGEFSDDLHPFNNSRSTNSNFIKGNVYVAKNPC 60

QY 760 LHPGDIVRLKAVNVRALHHVDCVVPQKGRPHNECSGSDLDGDIYFVCWQDMIPPR 819  
DB 61 LHPGDIVRLKAVNVRALHHVDCVVPQKGRPHNECSGSDLDGDIYFVCWQDMIPPR 120

QY 820 QVQPMEXPPAPSIQLDHDTVIEVEEYFTNYIVNDSLGIIANAHVVFADREPDMAMSDPC 879  
DB 121 QVQPMEXPPAPSIQLDHDTVIEVEEYFTNYIVNDSLGIIANAHVVFADREPDMAMSDPC 180

QY 880 KKLAEFSAIVDPKTPGPAEIPSQLRPKEYPDFMDKP 917  
DB 181 KKLAEFSAIVDPKTPGPAEIPSQLRPKEYPDFMDKP 218

```

RESULT 3
US-08-811-583-12
; Sequence 12, Application US/08811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wasseneeger, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RdRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; Zip: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,583
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-811-583-12

Query Match 1.5%; Score 17; DB 4; Length 17;
Best Local Similarity 100.0%; Pred. No. 8.6e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 331 EGYDGYLKGROPPKSPS 347
Db 1 EGYDGYLKGROPPKSPS 17
|||||
|||||

RESULT 4
US-08-811-583-10
; Sequence 10, Application US/08811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wasseneeger, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RdRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA

```

LENGTH: 14 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-811-583-11

Query Match 1.38; Score 14; DB 4; Length 14;  
Best Local Similarity 100.0%; Pred. No. 7.5e-07;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1007 ASKTFDRKDAEAI 1020  
Db 1 ASKTFDRKDAEAI 14

## RESULT 6

US-08-811-583-13  
Sequence 13, Application US/08811583  
Patent No. 6218142  
GENERAL INFORMATION:  
APPLICANT: Wassenecker, Michael  
APPLICANT: Riedel, Leonhard  
APPLICANT: Schiebel, Winfried  
APPLICANT: Sanger, Heinz  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
POLYPEPTIDES HAVING THE ENZYMAIC ACTIVITY OF AN  
RNA-DIRECTED RNA POLYMERASE (RGRP)  
TITLE OF INVENTION: RNP-DIRECTED RNA POLYMERASE (RGRP)  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/811,583

APPLICATION NUMBER: US/08/811,583

FILING DATE: 05-MAR-1997

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Haley, James F.

REGISTRATION NUMBER: 27,794

REFERENCE/DOCKET NUMBER: MPG-1

TELEPHONE: 212-596-9000

TELEFAX: 212-596-9090

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 13 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-811-583-13

Query Match 0.98; Score 10; DB 4; Length 13;  
Best Local Similarity 100.0%; Pred. No. 0.0073;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 784 VFPQKGRPH 793  
Db 1 VFPQKGRPH 10

## RESULT 7

US-08-949-202-2  
Sequence 2, Application US/08949202

Patent No. 6060587  
GENERAL INFORMATION:  
APPLICANT: Weiner, David B  
APPLICANT: Ayyavoo, Velupandi  
APPLICANT: Mahalingam, Sundarasamy  
APPLICANT: Patel, Mamata  
TITLE OF INVENTION: CELLULAR RECEPTOR FOR HIV-1 VPR ESSENTIAL FOR  
G2/M PHASE TRANSITION OF THE CELL CYCLE  
TITLE OF INVENTION: G2/M PHASE TRANSITION OF THE CELL CYCLE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6060587rls  
STREET: One Liberty Place 46th floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19103

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Windows NT  
SOFTWARE: Wordperfect 6.0  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/949,202

FILING DATE:

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Deluca, Mark

REGISTRATION NUMBER: 33,229

REFERENCE/DOCKET NUMBER: UPAP-0222

TELEPHONE: 215-568-3100

TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 341 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-949-202-2

Query Match 0.7%; Score 8; DB 3; Length 341;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 FFFFLSSG 137

Db 26 FFFFLSSG 33

## RESULT 8

US-08-949-202-4

Sequence 4, Application US/08949202

Patent No. 6060587

GENERAL INFORMATION:

APPLICANT: Weiner, David B

APPLICANT: Ayyavoo, Velupandi

APPLICANT: Mahalingam, Sundarasamy

APPLICANT: Patel, Mamata

TITLE OF INVENTION: CELLULAR RECEPTOR FOR HIV-1 VPR ESSENTIAL FOR

G2/M PHASE TRANSITION OF THE CELL CYCLE

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6060587rls

STREET: One Liberty Place 46th floor

CITY: Philadelphia

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows NT

SOFTWARE: Wordperfect 6.0



;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/949,202  
;; FILING DATE:  
;; CLASSIFICATION: 536  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Deluca, Mark  
;; REGISTRATION NUMBER: 33,229  
;; REFERENCE/DOCKET NUMBER: UPAP-0222  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 215-568-3100  
;; TELEFAX: 215-568-3439  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 341 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-949-202-4

Query Match 0.7%; Score 8; DB 3; Length 341;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 FFFFLSSG 137  
| | | | | | | |  
Db 26 FFFFLSSG 33

RESULT 9  
US-09-418-175-2  
; Sequence 2, Application US/09418175  
; Patent No. 6172201  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, David B  
; APPLICANT: Ayyavoo, Velpandi  
; APPLICANT: Mahalingam, Sundarasamy  
; APPLICANT: Patel, Mamata  
; TITLE OF INVENTION: CELLULAR RECEPTOR FOR HIV-1 VPR ESSENTIAL FOR  
; TITLE OF INVENTION: G2/M PHASE TRANSITION OF THE CELL CYCLE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6172201iris  
; STREET: One Liberty Place 46th floor  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: Windows NT  
; SOFTWARE: Wordperfect 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/418,175  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/949,202  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Deluca, Mark  
; REGISTRATION NUMBER: 33,229  
; REFERENCE/DOCKET NUMBER: UPAP-0222  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-568-3100  
; TELEFAX: 215-568-3439  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 341 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-418-175-2

Query Match 0.7%; Score 8; DB 4; Length 341;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 130 FFFFLSSG 137  
| | | | | | | |  
Db 26 FFFFLSSG 33

RESULT 10  
US-09-418-175-4  
; Sequence 4, Application US/09418175  
; Patent No. 6172201  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, David B  
; APPLICANT: Ayyavoo, Velpandi  
; APPLICANT: Mahalingam, Sundarasamy  
; APPLICANT: Patel, Mamata  
; TITLE OF INVENTION: CELLULAR RECEPTOR FOR HIV-1 VPR ESSENTIAL FOR  
; TITLE OF INVENTION: G2/M PHASE TRANSITION OF THE CELL CYCLE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6172201iris  
; STREET: One Liberty Place 46th floor  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: Windows NT  
; SOFTWARE: Wordperfect 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/418,175  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/949,202  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Deluca, Mark  
; REGISTRATION NUMBER: 33,229  
; REFERENCE/DOCKET NUMBER: UPAP-0222  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-568-3100  
; TELEFAX: 215-568-3439  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 341 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-418-175-4

Query Match 0.7%; Score 8; DB 4; Length 341;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 FFFFLSSG 137  
| | | | | | | |  
Db 26 FFFFLSSG 33

RESULT 11  
US-09-529-245-2  
; Sequence 2, Application US/09529245  
; Patent No. 6448078  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, David B  
; APPLICANT: Ayyavoo, Velpandi  
; APPLICANT: Mahalingam, Sundarasamy

```

; APPLICANT: Patel, Mamata
; TITLE OF INVENTION: Cellular Receptor For HIV-1 VPR Essential for G2/M
; FILE REFERENCE: UPAP0376
; CURRENT APPLICATION NUMBER: US/09/529,245
; CURRENT FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: 08/949,202
; PRIOR FILING DATE: 1997-10-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-529-245-2

Query Match      0.7%; Score 8; DB 4; Length 341;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 FFFFLSSG 137
Db 26 FFFFLSSG 33

RESULT 12
US-09-529-245-4
; Sequence 4, Application US/09529245
; Patent No. 6448078
; GENERAL INFORMATION:
; APPLICANT: Weiner, David B
; APPLICANT: Ayayoco, Velandí
; APPLICANT: Mahalingam, Sundarasamy
; APPLICANT: Patel, Mamata
; TITLE OF INVENTION: Cellular Receptor For HIV-1 VPR Essential for G2/M
; FILE REFERENCE: UPAP0376
; CURRENT APPLICATION NUMBER: US/09/529,245
; CURRENT FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: 08/949,202
; PRIOR FILING DATE: 1997-10-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-529-245-4

Query Match      0.7%; Score 8; DB 4; Length 341;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 FFFFLSSG 137
Db 26 FFFFLSSG 33

RESULT 13
US-09-457-040B-17
; Sequence 17, Application US/09457040B
; Patent No. 6387641
; GENERAL INFORMATION:
; APPLICANT: Vertex Pharmaceuticals Incorporated
; APPLICANT: Bellon, Steve
; TITLE OF INVENTION: Crystallized P38 Complexes
; FILE REFERENCE: VPI/98-14
; CURRENT APPLICATION NUMBER: US/09/457,040B
; CURRENT FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 506

```

```

; TYPE: PRT
; ORGANISM: Yeast
US-09-457-040B-17

Query Match      0.7%; Score 8; DB 4; Length 506;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 923 ISERVICK 930
Db 314 ISERVICK 321

RESULT 14
US-09-134-001C-3012
; Sequence 3012, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3012
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3012

Query Match      0.6%; Score 7; DB 4; Length 149;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 606 LITLLST 612
Db 18 LITLLST 24

RESULT 15
US-09-134-001C-3132
; Sequence 3132, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3132
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3132

Query Match      0.6%; Score 7; DB 4; Length 154;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 934 KVKDKAP 940
Db 51 KVKDKAP 57

```

RESULT 16  
US-09-455-960-7  
; Sequence 7, Application US/09455960  
; Patent No. 6361776  
; GENERAL INFORMATION:  
; APPLICANT: Delcayre, Alain  
; TITLE OF INVENTION: Compositions isolated from *M. vaccae* and  
; TITLE OF INVENTION: their use in modulation of immune responses.  
; FILE REFERENCE: 11000.1047  
; CURRENT APPLICATION NUMBER: US/09/455-960  
; CURRENT FILING DATE: 1999-12-06  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 7  
; LENGTH: 165  
; TYPE: PRT  
; ORGANISM: *Mycobacterium vaccae*  
US-09-455-960-7

Query Match 0.6%; Score 7; DB 4; Length 165;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 VASKGGL 545  
Db 34 VASKGGL 40

RESULT 17  
US-08-297-431B-31  
; Sequence 31, Application US/08297431B  
; Patent No. 6136605  
; GENERAL INFORMATION:  
; APPLICANT: Fahl, William E  
; APPLICANT: Gulick, Andrew M  
; APPLICANT: Manoharan, T Herbert  
; APPLICANT: Puchalski, Ralph B  
; APPLICANT: Kramer, Katharine  
; APPLICANT: Wasserman Wyeth W  
; TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saul, Ewing, Remick & Saul, LLP  
; STREET: 1500 Market Street, 38th Floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: US  
; ZIP: 19102-2186  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/297,431B  
; FILING DATE: August 26, 1994  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reed, Janet E.  
; REGISTRATION NUMBER: 36252  
; REFERENCE/DOCKET NUMBER: WARF F039  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-972-2292  
; TELEFAX: 215-972-8386  
; INFORMATION FOR SEQ ID NO: 31:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 220 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-297-431B-31

Query Match 0.6%; Score 7; DB 4; Length 220;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095  
Db 124 KDKARNR 130

RESULT 18  
US-08-297-431B-2  
; Sequence 2, Application US/08297431B  
; Patent No. 6136605  
; GENERAL INFORMATION:  
; APPLICANT: Fahl, William E  
; APPLICANT: Gulick, Andrew M  
; APPLICANT: Manoharan, T Herbert  
; APPLICANT: Puchalski, Ralph B  
; APPLICANT: Kramer, Katharine  
; APPLICANT: Wasserman Wyeth W  
; TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saul, Ewing, Remick & Saul, LLP  
; STREET: 1500 Market Street, 38th Floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: US  
; ZIP: 19102-2186  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/297,431B  
; FILING DATE: August 26, 1994  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reed, Janet E.  
; REGISTRATION NUMBER: 36252  
; REFERENCE/DOCKET NUMBER: WARF F039  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-972-8386  
; TELEFAX: 215-972-2292  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 221 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-297-431B-2

Query Match 0.6%; Score 7; DB 4; Length 221;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095  
Db 125 KDKARNR 131

RESULT 19  
US-08-297-431B-4  
; Sequence 4, Application US/08297431B  
; Patent No. 6136605  
; GENERAL INFORMATION:  
; APPLICANT: Fahl, William E  
; APPLICANT: Gulick, Andrew M  
; APPLICANT: Manoharan, T Herbert  
; APPLICANT: Puchalski, Ralph B  
; APPLICANT: Kramer, Katharine

```

; APPLICANT: Wasserman Wyeth W
; TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saul, Ewing, Remick & Saul, LLP
; STREET: 1500 Market Street, 38th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19102-2186
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297,431B
; FILING DATE: August 26, 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36252
; REFERENCE/DOCKET NUMBER: WARF F039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-972-8386
; TELEFAX: 215-972-2292
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-297-431B-4

Query Match 0.6%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
Db 125 KDKARNR 131

RESULT 20
US-08-297-431B-6
; Sequence 6, Application US/08297431B
; Patent No. 6136605
; GENERAL INFORMATION:
; APPLICANT: Fahl, William E.
; APPLICANT: Gulick, Andrew M.
; APPLICANT: Manoharan, T Herbert
; APPLICANT: Puchalski, Ralph B.
; APPLICANT: Kramer, Katharine
; APPLICANT: Wasserman Wyeth W
; TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saul, Ewing, Remick & Saul, LLP
; STREET: 1500 Market Street, 38th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19102-2186
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297,431B
; FILING DATE: August 26, 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36252
; REFERENCE/DOCKET NUMBER: WARF F039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-972-8386
; TELEFAX: 215-972-2292
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-297-431B-8

Query Match 0.6%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
Db 125 KDKARNR 131

RESULT 21
US-08-297-431B-8
; Sequence 8, Application US/08297431B
; Patent No. 6136605
; GENERAL INFORMATION:
; APPLICANT: Fahl, William E.
; APPLICANT: Gulick, Andrew M.
; APPLICANT: Manoharan, T Herbert
; APPLICANT: Puchalski, Ralph B.
; APPLICANT: Kramer, Katharine
; APPLICANT: Wasserman Wyeth W
; TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saul, Ewing, Remick & Saul, LLP
; STREET: 1500 Market Street, 38th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19102-2186
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297,431B
; FILING DATE: August 26, 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36252
; REFERENCE/DOCKET NUMBER: WARF F039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-972-8386
; TELEFAX: 215-972-2292
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-297-431B-8

Query Match 0.6%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
Db 125 KDKARNR 131
```

```

; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36252
; REFERENCE/DOCKET NUMBER: WARF F039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-972-8386
; TELEFAX: 215-972-2292
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-297-431B-6

Query Match 0.6%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
Db 125 KDKARNR 131

RESULT 21
US-08-297-431B-8
; Sequence 8, Application US/08297431B
; Patent No. 6136605
; GENERAL INFORMATION:
; APPLICANT: Fahl, William E.
; APPLICANT: Gulick, Andrew M.
; APPLICANT: Manoharan, T Herbert
; APPLICANT: Puchalski, Ralph B.
; APPLICANT: Kramer, Katharine
; APPLICANT: Wasserman Wyeth W
; TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saul, Ewing, Remick & Saul, LLP
; STREET: 1500 Market Street, 38th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19102-2186
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297,431B
; FILING DATE: August 26, 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36252
; REFERENCE/DOCKET NUMBER: WARF F039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-972-8386
; TELEFAX: 215-972-2292
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-297-431B-8

Query Match 0.6%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095
Db 125 KDKARNR 131
```

Db 125 KDKARNR 131

## RESULT 22

US-08-297-431B-10  
; Sequence 10, Application US/08297431B  
; Patent No. 6136605  
; GENERAL INFORMATION:  
; APPLICANT: Fahl, William E  
; APPLICANT: Gulick, Andrew M  
; APPLICANT: Manoharan, T Herbert  
; APPLICANT: Puchalski, Ralph B  
; APPLICANT: Kramer, Katharine  
; APPLICANT: Wasserman Wyeth W  
; TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saul, Ewing, Remick & Saul, LLP  
; STREET: 1500 Market Street, 38th Floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: US  
; ZIP: 19102-2186  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE: August 26, 1994  
; APPLICATION NUMBER: US/08/297,431B  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reed, Janet E.  
; REGISTRATION NUMBER: 36252  
; REFERENCE/DOCKET NUMBER: WARF F039  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-972-8386  
; TELEFAX: 215-972-2292  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 221 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-297-431B-10

Query Match 0.6%; Score 7; DB 4; Length 221;  
Best Local Similarity 100.0%; Pred. No. le+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1089 KDKARNR 1095

|||||||

Db 125 KDKARNR 131

## RESULT 23

US-08-297-431B-12  
; Sequence 12, Application US/08297431B  
; Patent No. 6136605  
; GENERAL INFORMATION:  
; APPLICANT: Fahl, William E  
; APPLICANT: Gulick, Andrew M  
; APPLICANT: Manoharan, T Herbert  
; APPLICANT: Puchalski, Ralph B  
; APPLICANT: Kramer, Katharine  
; APPLICANT: Wasserman Wyeth W  
; TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saul, Ewing, Remick & Saul, LLP  
; STREET: 1500 Market Street, 38th Floor  
; CITY: Philadelphia

STATE: PA  
COUNTRY: US  
ZIP: 19102-2186  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: August 26, 1994  
APPLICATION NUMBER: US/08/297,431B  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Reed, Janet E.  
REGISTRATION NUMBER: 36252  
REFERENCE/DOCKET NUMBER: WARF F039  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-972-8386  
TELEFAX: 215-972-2292  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 221 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-297-431B-12

Query Match 0.6%; Score 7; DB 4; Length 221;  
Best Local Similarity 100.0%; Pred. No. le+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1089 KDKARNR 1095

|||||||

Db 125 KDKARNR 131

## RESULT 24

US-08-297-431B-14  
; Sequence 14, Application US/08297431B  
; Patent No. 6136605  
; GENERAL INFORMATION:  
; APPLICANT: Fahl, William E  
; APPLICANT: Gulick, Andrew M  
; APPLICANT: Manoharan, T Herbert  
; APPLICANT: Puchalski, Ralph B  
; APPLICANT: Kramer, Katharine  
; APPLICANT: Wasserman Wyeth W  
; TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saul, Ewing, Remick & Saul, LLP  
; STREET: 1500 Market Street, 38th Floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: US  
; ZIP: 19102-2186  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE: August 26, 1994  
; APPLICATION NUMBER: US/08/297,431B  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reed, Janet E.  
; REGISTRATION NUMBER: 36252  
; REFERENCE/DOCKET NUMBER: WARF F039  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-972-8386  
; TELEFAX: 215-972-2292  
; INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:  
LENGTH: 221 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
US-08-297-431B-14

Query Match 0.6%; Score 7; DB 4; Length 221;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095  
Db 125 KDKARNR 131

RESULT 25  
US-08-297-431B-16  
Sequence 16, Application US/08297431B  
Patent No. 6136605  
GENERAL INFORMATION:  
APPLICANT: Fahl, William E  
APPLICANT: Gulick, Andrew M  
APPLICANT: Manoharan, T Herbert  
APPLICANT: Puchalski, Ralph B  
APPLICANT: Kramer, Katharine  
APPLICANT: Wasserman Wyeth W  
TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saul, Ewing, Remick & Saul, LLP  
STREET: 1500 Market Street, 38th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19102-2186  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/297,431B  
FILING DATE: August 26, 1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Reed, Janet E.  
REGISTRATION NUMBER: 36252  
REFERENCE/DOCKET NUMBER: WARF F039  
TELEPHONE: 215-972-8386  
TELEFAX: 215-972-2292  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 221 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
US-08-297-431B-16

Query Match 0.6%; Score 7; DB 4; Length 221;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095  
Db 125 KDKARNR 131

RESULT 26  
US-08-297-431B-18  
Sequence 18, Application US/08297431B  
Patent No. 6136605

GENERAL INFORMATION:  
APPLICANT: Fahl, William E  
APPLICANT: Gulick, Andrew M  
APPLICANT: Manoharan, T Herbert  
APPLICANT: Puchalski, Ralph B  
APPLICANT: Kramer, Katharine  
APPLICANT: Wasserman Wyeth W  
TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saul, Ewing, Remick & Saul, LLP  
STREET: 1500 Market Street, 38th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19102-2186  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/297,431B  
FILING DATE: August 26, 1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Reed, Janet E.  
REGISTRATION NUMBER: 36252  
REFERENCE/DOCKET NUMBER: WARF F039  
TELEPHONE: 215-972-8386  
TELEFAX: 215-972-2292  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 221 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
US-08-297-431B-18

Query Match 0.6%; Score 7; DB 4; Length 221;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095  
Db 125 KDKARNR 131

RESULT 27  
US-08-297-431B-20  
Sequence 20, Application US/08297431B  
Patent No. 6136605  
GENERAL INFORMATION:  
APPLICANT: Fahl, William E  
APPLICANT: Gulick, Andrew M  
APPLICANT: Manoharan, T Herbert  
APPLICANT: Puchalski, Ralph B  
APPLICANT: Kramer, Katharine  
APPLICANT: Wasserman Wyeth W  
TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saul, Ewing, Remick & Saul, LLP  
STREET: 1500 Market Street, 38th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19102-2186  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/297,431B  
;; FILING DATE: August 26, 1994  
;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Reed, Janet E.  
;; REGISTRATION NUMBER: 36252  
;; REFERENCE/DOCKET NUMBER: WARF F039  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 215-972-8386  
;; TELEFAX: 215-972-2292  
;; INFORMATION FOR SEQ ID NO: 20:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 221 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-297-431B-20

Query Match 0.6%; Score 7; DB 4; Length 221;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1089 KDKARNR 1095  
|||||||  
DB 125 KDKARNR 131

RESULT 28  
US-08-565-386-17  
; Sequence 17, Application US/08565386  
; Patent No. 5741697  
; GENERAL INFORMATION:  
; APPLICANT: Bavoll, Patrik M.  
; APPLICANT: Hsia, Ru-ching  
; TITLE OF INVENTION: BACTERIOPHAGE OF CHLAMYDIA PSITTACI  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: Clinton Square, P.O. Box 1051  
; CITY: Roches  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/565,386  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Timian, Susan J.  
; REGISTRATION NUMBER: 34,103  
; REFERENCE/DOCKET NUMBER: 176/60040  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 716-263-1636  
; TELEFAX: 716-263-1600  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 320 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-565-386-17

Query Match 0.6%; Score 7; DB 1; Length 320;  
Best Local Similarity 100.0%; Pred. No. 1.4e-02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 849 NYIVNDS 855  
|||||||

DB 36 NYIVNDS 42

RESULT 29  
US-08-576-626A-33  
; Sequence 33, Application US/08576626A  
; Patent No. 5998194  
; GENERAL INFORMATION:  
; APPLICANT: Summers, R.G.  
; APPLICANT: Katz, L.  
; APPLICANT: Donadio, S.  
; APPLICANT: Staver, M.J.  
; TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR  
; TITLE OF INVENTION: BIOSYNTHESIS GENES  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/576,626A  
; FILING DATE: 21-DEC-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dianne Casuto  
; REGISTRATION NUMBER: P-40,943  
; REFERENCE/DOCKET NUMBER: 5857.US.O1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (847) 938-3137  
; TELEFAX: (847) 938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 33:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 322 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5998194e  
US-08-576-626A-33  
Query Match 0.6%; Score 7; DB 2; Length 322;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 604 RQLITLL 610  
|||||||  
DB 9 RQLITLL 15  
RESULT 30  
US-08-791-115B-27  
; Sequence 27, Application US/08791115B  
; Patent No. 6262242  
; GENERAL INFORMATION:  
; APPLICANT: Steck, Peter  
; APPLICANT: Pershouse, Mark A.  
; APPLICANT: Jasser, Samar  
; APPLICANT: Yung, W.K. Alfred  
; APPLICANT: Tavtigian, Sean V.  
; TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10Q23.3  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.  
STREET: 555 Thirteenth Street, N.W., Suite 701-E  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 22204  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/791.115B  
APPLICATION NUMBER: US/08/791.115B  
FILING DATE: 30-JAN-1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ihnen, Jeffrey L.  
REGISTRATION NUMBER: 38,957  
REFERENCE/DOCKET NUMBER: 2318-134.A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-683-7031  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 394 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-791-115B-27

Query Match 0.6%; Score 7; DB 4; Length 394;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 639 LKAEAL 645  
DB 146 LKAEAL 152

RESULT 31  
US-08-791-115B-1  
; Sequence 1, Application US/08791115B  
; Patent No. 6262242  
; GENERAL INFORMATION:  
; APPLICANT: Steck, Peter  
; APPLICANT: Pershouse, Mark A.  
; APPLICANT: Jasser, Samar  
; APPLICANT: Yung, W.K. Alfred  
; APPLICANT: Tavtigian, Sean V.  
; TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10Q23.3  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.  
; STREET: 555 Thirteenth Street, N.W., Suite 701-E  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 22204  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/791.115B  
; FILING DATE: 30-JAN-1997  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.  
; REGISTRATION NUMBER: 38,957  
; REFERENCE/DOCKET NUMBER: 2318-134.A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-683-6040

TELEFAX: 202-683-7031  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 403 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-791-115B-1  
Query Match 0.6%; Score 7; DB 4; Length 403;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 639 LKAEAL 645  
DB 146 LKAEAL 152

RESULT 32  
US-08-791-115B-23  
; Sequence 23, Application US/08791115B  
; Patent No. 6262242  
; GENERAL INFORMATION:  
; APPLICANT: Steck, Peter  
; APPLICANT: Pershouse, Mark A.  
; APPLICANT: Jasser, Samar  
; APPLICANT: Yung, W.K. Alfred  
; APPLICANT: Tavtigian, Sean V.  
; TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10Q23.3  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.  
; STREET: 555 Thirteenth Street, N.W., Suite 701-E  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 22204  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/791.115B  
; FILING DATE: 30-JAN-1997  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.  
; REGISTRATION NUMBER: 38,957  
; REFERENCE/DOCKET NUMBER: 2318-134.A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-683-6040  
; TELEFAX: 202-683-7031  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 403 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-791-115B-23

Query Match 0.6%; Score 7; DB 4; Length 403;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 639 LKAEAL 645  
DB 146 LKAEAL 152

RESULT 33  
US-08-791-115B-25  
; Sequence 25, Application US/08791115B



```
; Patent No. 6262242
; GENERAL INFORMATION:
; APPLICANT: Steck, Peter
; APPLICANT: Pershouse, Mark A.
; APPLICANT: Jasser, Samar
; APPLICANT: Yung, W.K. Alfred
; APPLICANT: Tavtigian, Sean V.
; TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10Q23.3
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
; STREET: 555 Thirteenth Street, N.W., Suite 701-E
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/791,115B
; FILING DATE: 30-JAN-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Innen, Jeffrey L.
; REGISTRATION NUMBER: 38,957
; REFERENCE/DOCKET NUMBER: 2318-134.A
; TELEPHONE: 202-683-6040
; TELEFAX: 202-683-7031
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-791-115B-25

Query Match 0.6%; Score 7; DB 4; Length 403;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 639 LKAQAL 645
Db 146 LKAQAL 152
|||||

RESULT 34
US-08-791-115B-7
; Sequence 7, Application US/08791115B
; Patent No. 6262242
; GENERAL INFORMATION:
; APPLICANT: Steck, Peter
; APPLICANT: Pershouse, Mark A.
; APPLICANT: Jasser, Samar
; APPLICANT: Yung, W.K. Alfred
; APPLICANT: Tavtigian, Sean V.
; TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10Q23.3
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
; STREET: 555 Thirteenth Street, N.W., Suite 701-E
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/791,115B
; FILING DATE: 30-JAN-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Innen, Jeffrey L.
; REGISTRATION NUMBER: 38,957
; REFERENCE/DOCKET NUMBER: 2318-134.A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-683-6040
; TELEFAX: 202-683-7031
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 430 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-791-115B-7

Query Match 0.6%; Score 7; DB 4; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 639 LKAQAL 645
Db 182 LKAQAL 188
|||||

RESULT 35
US-08-933-821-6
; Sequence 6, Application US/08933821
; Patent No. 5972338
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: Tie Ligands
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,821
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 470 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-933-821-6

Query Match 0.6%; Score 7; DB 2; Length 470;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1024 VRAIRKE 1030
|||||
```

Db 93 VRLRKE 99

## RESULT 36

US-08-960-507-6

; Sequence 6, Application US/08960507

; Patent No. 6057435

; GENERAL INFORMATION:

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; TITLE OF INVENTION: Tie Ligands

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 1 DNA Way

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: WinPatIn (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/960,507

; FILING DATE:

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Dreger, Ginger R.

; REGISTRATION NUMBER: 33,055

; REFERENCE/DOCKET NUMBER: P1130p1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650/225-3216

; TELEFAX: 650/952-9881

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 470 amino acids

; TYPE: Amino Acid

; TOPOLOGY: Linear

US-08-960-507-6

Query Match 0.6%; Score 7; DB 3; Length 470;

Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1024 VRLRKE 1030

Db 93 VRLRKE 99

## RESULT 37

US-09-136-828-6

; Sequence 6, Application US/09136828

; Patent No. 6350450

; GENERAL INFORMATION:

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; TITLE OF INVENTION: Tie Ligands

; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 1 DNA Way

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: WinPatIn (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,828

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Dreger, Ginger R.

; REGISTRATION NUMBER: 33,055

; REFERENCE/DOCKET NUMBER: P1130R1A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650/225-3216

; TELEFAX: 650/952-9881

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 470 amino acids

; TYPE: Amino Acid

; TOPOLOGY: Linear

US-09-136-828-6

Query Match 0.6%; Score 7; DB 4; Length 470;

Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1024 VRLRKE 1030

Db 93 VRLRKE 99

## RESULT 38

US-09-332-928A-6

; Sequence 6, Application US/09332928A

; Patent No. 6368853

; GENERAL INFORMATION:

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; TITLE OF INVENTION: Tie Ligands

; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 1 DNA Way

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: WinPatIn (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/332,928A

; FILING DATE: 14-Jun-1999

; CLASSIFICATION: &lt;Unknown&gt;

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/933,821

; FILING DATE: &lt;Unknown&gt;

; ATTORNEY/AGENT INFORMATION:

; NAME: Dreger, Ginger R.

; REGISTRATION NUMBER: 33,055

; REFERENCE/DOCKET NUMBER: P1130

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650/225-3216

; TELEFAX: 650/952-9881

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 470 amino acids

; TYPE: Amino Acid

; TOPOLOGY: Linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-332-928A-6

Query Match 0.6%; Score 7; DB 4; Length 470;

Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1024 VRLRKE 1030  
 Db 93 VRLRKE 99

## RESULT 39

US-09-136-801-6  
 ; Sequence 6, Application US/09136801  
 ; Patent No. 6413770  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin  
 ; APPLICANT: Hillan, Kenneth  
 ; APPLICANT: Botstein, David  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Roy, Margaret  
 ; APPLICANT: Ferrara, Napoleone  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Schwall, Ralph  
 ; TITLE OF INVENTION: Tie Ligand Homologues  
 ; NUMBER OF SEQUENCES: 35  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 1 DNA Way  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080

COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: WinPatIn (Genentech)

CURRENT APPLICATION DATA: US/09/136, 801

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Dreger, Ginger R.

REGISTRATION NUMBER: 33,055

REFERENCE/DOCKET NUMBER: P1130P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/225-3216

TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 470 amino acids

TYPE: Amino Acid

TOPOLOGY: Linear

US-09-136-801-6

Query Match 0.6%; Score 7; DB 4; Length 470;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1024 VRLRKE 1030  
 Db 93 VRLRKE 99

## RESULT 40

US-09-332-929-6  
 ; Sequence 6, Application US/09332929  
 ; Patent No. 6420542  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; TITLE OF INVENTION: Tie Ligands  
 ; NUMBER OF SEQUENCES: 17  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 1 DNA Way  
 ; CITY: South San Francisco

STATE: California  
 COUNTRY: USA  
 ZIP: 94080  
 COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: WinPatIn (Genentech)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/332, 929  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/933, 821  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Dreger, Ginger R.  
 ; REGISTRATION NUMBER: 33,055  
 ; REFERENCE/DOCKET NUMBER: P1130  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 650/225-3216  
 ; TELEFAX: 650/952-9881  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 470 amino acids  
 ; TYPE: Amino Acid  
 ; TOPOLOGY: Linear  
 ; US-09-332-929-6

Query Match 0.6%; Score 7; DB 4; Length 470;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1024 VRLRKE 1030  
 Db 93 VRLRKE 99

## RESULT 41

US-08-968-563-15  
 ; Sequence 15, Application US/08968563  
 ; Patent No. 6013494  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CHARLES E. NAKAMURA  
 ; APPLICANT: ANTHONY A. GATENBY  
 ; APPLICANT: AMY (KUANG-HUA) HSU  
 ; APPLICANT: RICHARD D. LA REAU  
 ; APPLICANT: SHARON L. HAYNIE  
 ; APPLICANT: MARIA DIAZ-TORRES  
 ; APPLICANT: DONALD E. TRIMBUR  
 ; APPLICANT: GREGORY M. WHITED  
 ; APPLICANT: VASANTHA NAGARAJAN  
 ; APPLICANT: MARK S. PAYNE  
 ; APPLICANT: STEPHEN K. PICATAGGIO  
 ; APPLICANT: RAMESCH V. NAIR  
 ; TITLE OF INVENTION: METHOD FOR THE RECOMBINANT  
 ; TITLE OF INVENTION: PRODUCTION OF 1,3-PROPANEDIOL  
 ; NUMBER OF SEQUENCES: 49  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
 ; STREET: 1007 MARKET STREET  
 ; CITY: WILMINGTON  
 ; STATE: DELAWARE  
 ; COUNTRY: U.S.A.  
 ; ZIP: 19898  
 ; ADDRESSEE: GENECOR INTERNATIONAL, INC.  
 ; STREET: 4 CAMBRIDGE PLACE  
 ; STREET: 1870 SOUTH WINTON ROAD  
 ; CITY: ROCHESTER  
 ; STATE: NEW YORK  
 ; COUNTRY: U.S.A.  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.50 INCH DISKETTE

COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95  
SOFTWARE: MICROSOFT WORD VERSION 7.0A  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/968,563  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/030,601  
FILING DATE: NOVEMBER 13, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: FLOYD, LINDA AXAMETHY  
REGISTRATION NUMBER: 33,692  
REFERENCE/DOCKET NUMBER: CR-9982  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-892-8112  
TELEFAX: 302-773-0164  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 501 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
ORIGINAL SOURCE: GLPD  
ORGANISM: GLPD  
US-08-968-563-15

Query Match 0.6%; Score 7; DB 3; Length 501;  
Best Local Similarity 100.0%; Pred. No. 2.2e-02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 309 EHALEKL 315  
Db 363 EHALEKL 369

RESULT 42  
US-08-969-683A-15  
Sequence 15, Application US/08969683A  
Patent No. 6136576  
GENERAL INFORMATION:  
APPLICANT: GENENCOR INTERNATIONAL, INC.  
TITLE OF INVENTION: METHOD FOR THE RECOMBINANT  
PRODUCTION OF 1,3 PROPANEDIOL  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genencor International, Inc.  
STREET: 4 Cambridge Place  
STREET: 1870 South Winton road  
CITY: Rochester  
STATE: NY  
COUNTRY: U.S.A  
ZIP: 14618  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/969,683A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/20873  
FILING DATE: 13-NOV-1997  
APPLICATION NUMBER: 60/030,601  
FILING DATE: 13-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Graister, Dedra  
REGISTRATION NUMBER: 33,888  
REFERENCE/DOCKET NUMBER: GC 369-2  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-864-7620  
TELEFAX: 650-845-6504  
TELEX:  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 501 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
ORIGINAL SOURCE: GLPD  
ORGANISM: GLPD  
US-08-969-683A-15

Query Match 0.6%; Score 7; DB 4; Length 501;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 309 EHALEKL 315  
Db 363 EHALEKL 369

RESULT 43  
US-09-297-928-11  
Sequence 11, Application US/09297928  
Patent No. 6358716  
GENERAL INFORMATION:  
APPLICANT: BULTHUIS, BEN A.  
GATENBY, ANTHONY A.  
HAYNIE, SHARON L.  
HSU, AMY K.  
LAREAU, RICHARD D.  
TITLE OF INVENTION: METHOD FOR THE PRODUCTION OF  
GLYCEROL BY RECOMBINANT  
ORGANISMS  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENENCOR INTERNATIONAL, INC.  
STREET: 4 CAMBRIDGE PLACE  
CITY: ROCHESTER  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 14618  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MICROSOFT WINDOWS 95  
SOFTWARE: MICROSOFT WORD VERSION 7.0A  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/297,928  
FILING DATE: 11-May-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/03602  
FILING DATE: NOVEMBER 13, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: FLOYD, LINDA AXAMETHY  
REGISTRATION NUMBER: 33,692  
REFERENCE/DOCKET NUMBER: CR-9981-P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-892-8112  
TELEFAX: 302-773-0164  
TELEX: 6717325  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 501 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 11:

US-09-297-928-11

Query Match 0.6%; Score 7; DB 4; Length 501;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 309 EHALEKL 315  
| | | | |  
DB 363 EHALEKL 369

RESULT 44

US-08-906-156A-12  
; Sequence 12, Application US/08906156A  
; Patent No. 6287854  
; GENERAL INFORMATION:  
; APPLICANT: SPURR, NIGEL K  
; APPLICANT: GRAY, IAN C  
; APPLICANT: STEWART, LORNA M  
; TITLE OF INVENTION: DIAGNOSIS OF SUSCEPTIBILITY TO CANCER  
; TITLE OF INVENTION: AND TREATMENT THEREOF  
; NUMBER OF SEQUENCES: 94  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERVE P.C.  
; STREET: 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22201  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/906,156A  
; FILING DATE: 05-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/042,655  
; FILING DATE: 02-APR-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/033,147  
; FILING DATE: 13-DEC-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/005,840  
; FILING DATE: 23-OCT-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/96GB/02588  
; FILING DATE: 22-OCT-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SADOFF, B. J.  
; REGISTRATION NUMBER: 36,663  
; REFERENCE/DOCKET NUMBER: 1090-14  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-816-4000  
; TELEFAX: 703-816-4100  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 565 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: translation of partial cDNA sequence

US-08-906-156A-12

Query Match 0.6%; Score 7; DB 4; Length 565;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 639 LKAQCAL 645  
| | | | |  
DB 308 LKAQCAL 314

RESULT 45

US-09-183-266A-49  
; Sequence 49, Application US/09183266A  
; Patent No. 6361954  
; GENERAL INFORMATION:  
; APPLICANT: Stillman, Bruce  
; APPLICANT: Williams, R. Sanders  
; APPLICANT: Mendez, Juan  
; TITLE OF INVENTION: ANTIBODIES THERETO AND DIAGNOSTIC APPLICATIONS THEREOF  
; TITLE OF INVENTION: DNA REPLICATION-REGULATING GENES,  
; FILE REFERENCE: CSHL96-01A3  
; CURRENT APPLICATION NUMBER: US/09/183,266A  
; CURRENT FILING DATE: 1998-10-30  
; PRIOR APPLICATION NUMBER: PCT/US97/07333  
; PRIOR FILING DATE: 1997-05-02  
; PRIOR APPLICATION NUMBER: 08/648,650  
; PRIOR FILING DATE: 1996-05-15  
; PRIOR APPLICATION NUMBER: 08/643,034  
; PRIOR FILING DATE: 1998-05-02  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 49  
; LENGTH: 577  
; TYPE: PRT  
; ORGANISM: S. cerevisiae  
US-09-183-266A-49

Query Match 0.6%; Score 7; DB 4; Length 577;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 572 KLSLRKS 578  
| | | | |  
DB 157 KLSLRKS 163

RESULT 46

US-08-791-115B-6  
; Sequence 6, Application US/08791115B  
; Patent No. 6262242  
; GENERAL INFORMATION:  
; APPLICANT: Steck, Peter  
; APPLICANT: Pershouse, Mark A.  
; APPLICANT: Jasser, Samar  
; APPLICANT: Yung, W.K. Alfred  
; APPLICANT: Tavtighian, Sean V.  
; TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10Q23.3  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.  
; STREET: 555 Thirteenth Street, N.W., Suite 701-E  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 22204  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/791,115B  
; FILING DATE: 30-JAN-1997  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.  
; REGISTRATION NUMBER: 38,957  
; REFERENCE/DOCKET NUMBER: 2318-134.A

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-683-6040  
; TELEFAX: 202-683-7031  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 645 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-791-115B-6

Query Match 0.6%; Score 7; DB 4; Length 645;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 639 LKAQEL 645  
Db 388 LKAQEL 394  
|||||

## RESULT 47

; Sequence 5, Application US/08791115B  
; Patent No. 626242  
; GENERAL INFORMATION:  
; APPLICANT: Steck, Peter  
; APPLICANT: Pershouse, Mark A.  
; APPLICANT: Jasser, Samar  
; APPLICANT: Yung, W.K. Alfred  
; APPLICANT: Tavligian, Sean V.  
; TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10Q23.3  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.  
; STREET: 555 Thirteenth Street, N.W., Suite 701-E  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 22204

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/791,115B  
; FILING DATE: 30-JAN-1997  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.  
; REGISTRATION NUMBER: 38,957  
; REFERENCE/DOCKET NUMBER: 2318-134.A

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-683-6040  
; TELEFAX: 202-683-7031  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 742 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-791-115B-5

Query Match 0.6%; Score 7; DB 4; Length 742;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 639 LKAQEL 645  
Db 485 LKAQEL 491  
|||||

## RESULT 48

; TITLE OF INVENTION: MAMMALIAN REGULATOR OF  
; NONSENSE-MEDIATED RNA DECAY

US-09-143-571-29  
; Sequence 29, Application US/09143571  
; Patent No. 633153  
; GENERAL INFORMATION:  
; APPLICANT: FISHEL, Richard  
; APPLICANT: GRADIA, Scott  
; APPLICANT: ACHARYA, Samir  
; TITLE OF INVENTION: NUCLEOTIDE MODULATION OF DNA MISMATCH RECOGNITION  
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR EFFECTING ADENINE  
; FILE REFERENCE: 9855-601  
; CURRENT APPLICATION NUMBER: US/09/143,571  
; CURRENT FILING DATE: 1998-08-28  
; EARLIER APPLICATION NUMBER: 60/093,935  
; EARLIER FILING DATE: 1998-07-23  
; EARLIER APPLICATION NUMBER: 60/066,977  
; EARLIER FILING DATE: 1997-11-28  
; EARLIER APPLICATION NUMBER: 60/057,136  
; EARLIER FILING DATE: 1997-08-28  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 29  
; LENGTH: 834  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-143-571-29

Query Match 0.6%; Score 7; DB 4; Length 834;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 728 EFSDDLH 734  
Db 366 EFSDDLH 372  
|||||

## RESULT 49

US-09-741-150-4  
; Sequence 4, Application US/097411150  
; Patent No. 6436689  
; GENERAL INFORMATION:  
; APPLICANT: GUEGLER, Karl et al  
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND  
; FILE REFERENCE: CLO000968  
; CURRENT APPLICATION NUMBER: US/09/741,150  
; CURRENT FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 884  
; TYPE: PRT  
; ORGANISM: Human  
US-09-741-150-4

Query Match 0.6%; Score 7; DB 4; Length 884;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 619 VLEQKOK 625  
Db 193 VLEQKOK 199  
|||||

## RESULT 50

US-08-724-354D-22  
; Sequence 22, Application US/08724354D  
; Patent No. 5994119  
; GENERAL INFORMATION:  
; APPLICANT: Dietz, Harry C.  
; TITLE OF INVENTION: MAMMALIAN REGULATOR OF  
; NONSENSE-MEDIATED RNA DECAY

; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/724,354D  
; FILING DATE: 01-OCT-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/016,482  
; FILING DATE: 29-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Halle, Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 07265/090001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-678-5070  
; TELEFAX: 619-678-5039  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 971 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-724-354D-22

Query Match 0.6%; Score 7; DB 2; Length 971;  
Best Local Similarity 100.0%; Pred.No. 4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 994 KTEAEIL 1000

Db 538 KTEAEIL 544

Search completed: November 6, 2002, 03:40:49  
Job time : 93 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 23:19:32 : Search time 4531 Seconds  
(without alignments)  
13335.988 Million cell updates/sec

Title: US-09-782-874-1

Perfect score: 3731

Sequence: 1 GAATATCTTACTTACTT.....AGTTTCATCTTCTCTCTAAA 3731

Scoring table: OLIGO\_NUC  
Gapop 60.0, Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database:

```

1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estrc:*
9: gb_estli:*
10: gb_estc2:*
11: gb_htc2:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	575	15.4	659	13	BI203182
2	546	14.6	618	10	BE435332
3	525	14.1	525	10	AW933692
4	421	11.3	421	9	AI896465
5	398	10.7	570	10	BE433846
6	358	9.6	537	10	AW932918

7	341	9.1	597	10	AW649811
8	316	8.5	367	12	BF112386
9	195	5.2	725	12	BG594667
10	141	3.8	521	14	BQ111608
11	141	3.8	610	12	BQ111609
12	121	3.2	434	14	BQ111609
13	110	2.9	403	12	BF052042
14	101	2.7	682	12	BG597875
15	98	2.6	773	14	BQ507474
16	95	2.5	443	10	AW623969
17	84	2.3	529	12	BG592782
18	81	2.2	393	12	BG889665
19	78	2.1	321	10	BE340733
20	71	1.9	550	9	AJ487433
21	57	1.5	340	14	BQ509665
22	24	0.6	541	14	BQ827639
23	23	0.6	348	17	BH258389
24	23	0.6	571	14	BQ488250
25	22	0.6	265	10	BR013867
26	22	0.6	308	9	AV045785
27	22	0.6	312	9	AV045898
28	22	0.6	332	9	AV210762
29	22	0.6	356	9	AV046870
30	22	0.6	580	10	AW644365
31	21	0.6	275	10	AV332932
32	21	0.6	282	9	AV207607
33	21	0.6	284	12	BG379160
34	21	0.6	323	10	BA446629
35	21	0.6	381	9	AV202914
36	21	0.6	426	10	BE581464
37	21	0.6	440	17	AQ498746
38	21	0.6	484	12	BG705217
39	21	0.6	500	13	BI693834
40	21	0.6	504	17	A2930706
41	21	0.6	521	13	BI038396
42	21	0.6	529	13	BI041199
43	21	0.6	630	17	AQ985038
44	21	0.6	637	17	A2866201
45	21	0.6	755	17	BH098015
46	21	0.6	765	17	BH608326
47	21	0.6	805	13	BG985828
48	21	0.6	808	17	BH558896
49	21	0.6	954	12	BF621359
50	21	0.6	1009	17	CNS0261X

#### ALIGNMENTS

RESULT 1  
BI203182

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BI203182 659 bp mRNA linear EST 11-JUL-2001  
EST521222 cTOS Lycopersicon esculentum cDNA clone cTOS1C7 5' end  
similar to putative RNA-directed RNA polymerase, mRNA sequence.  
BI203182  
BI203182.1 GI:14680906  
EST.  
tomato.  
Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 659)  
van der Hoeven, R., Sun, H., Bezzerides, J., Cho, J., Utterback, R.,  
Ronning, C. and Tanksley, S.  
Generation of ESTs from Tomato Suspension Cultures  
Unpublished (2001)  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>.



```

FEATURES          Location/Qualifiers
source            1..659
  /organism="Lycopersicon esculentum"
  /cultivar="TA496, E6203"
  /db_xref="taxon:4081"
  /clone="cFOS1C7"
  /clone_lib="cFOS"
  /tissue_type="suspension cultures"
  /lab_host="SOLR"
  /note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Suspension cultures of L.esculentum E6203 were grown
in Murashige and Skoog based medium, supplemented with 1%
coconut milk (filter sterilized and added after
autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).
Fresh medium was added every 7 days, and cultures were
grown at 25 C, with 12hrs of light and continuous
shaking."
BASE COUNT       210 a 107 c 153 g 189 t
ORIGIN
Query Match      15.4%; Score 575; DB 13; Length 659;
Best Local Similarity 100.0%; Pred. No. 8.8e-294;
Matches 575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3060 CATATGATGCTGATGATGAAGTTCATGATTTGAAGATTACATTCAGGAGCTTTTGACT 3119
DB 1 CATATGATGCTGATGATGAAGTTCATGATTTGAAGATTACATTCAGGAGCTTTTGACT 60
QY 3120 ACAAACCTGATATGACAAACAGCTGGTATTTAATGAGCTACTATGGCATATAAACAG 3179
DB 61 ACAAACCTGATATGACAAACAGCTGGTATTTAATGAGCTACTATGGCATATAAACAG 120
QY 3180 AGGCTGAATACTAGTGGTGGCATTTAAGAGCATCAAAACTTTTGCCGAGAAAAG 3239
DB 121 AGGCTGAATACTAGTGGTGGCATTTAAGAGCATCAAAACTTTTGCCGAGAAAAG 180
QY 3240 ATGCTGAGGCCATAGTGTCTGTGAGGGCTTGAGGAGGAGGCAAGAGCTGGTTCA 3299
DB 181 ATGCTGAGGCCATAGTGTCTGTGAGGGCTTGAGGAGGAGGCAAGAGCTGGTTCA 240
QY 3300 AGAGCGTAATGATATGATGACATGTTACCAAGGCTTCGGCTTGGTACCAGCTTACAT 3359
DB 241 AGAGCGTAATGATATGATGACATGTTACCAAGGCTTCGGCTTGGTACCAGCTTACAT 300
QY 3360 ATCATCTACATATGGGGTGGCTACATCAGGGTTGAAAGAGCTCATTTTCATTAGCT 3419
DB 301 ATCATCTACATATGGGGTGGCTACATCAGGGTTGAAAGAGCTCATTTTCATTAGCT 360
QY 3420 TTCCCTGGTGTGTTATGACACAGCTAATCCAGATTAAAGAGGACAAAGCACCTAACAGGC 3479
DB 361 TTCCCTGGTGTGTTATGACACAGCTAATCCAGATTAAAGAGGACAAAGCACCTAACAGGC 420
QY 3480 CAGTTCTCAACTGTCTCTCAGGGCTCAACTGAGTCACAGATTAGTTGAATGAG 3539
DB 421 CAGTTCTCAACTGTCTCTCAGGGCTCAACTGAGTCACAGATTAGTTGAATGAG 480
QY 3540 ATTCCAGTCGAGCGTTAAGCTGATATATATATATGATGATGAGGTGTGATCATAAGAAA 3599
DB 481 ATTCCAGTCGAGCGTTAAGCTGATATATATATATGATGATGAGGTGTGATCATAAGAAA 540
QY 3600 CTGTTATGCAATGTTGACTACCTTTTCTTTTAA 3634
DB 541 CTGTTATGCAATGTTGACTACCTTTTCTTTTAA 575

RESULT 2
BE435332
LOCUS
DEFINITION EST406410 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
ACCESSION BE435332
VERSION BE435332.1 GI:9433175
KEYWORDS EST.

```

tomato.

ORGANISM  
Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 618)

REFERENCE  
AUTHORS  
Alcala J., Vrebalov J., White R., van der Hoeven R.S., Holt, I.E.,  
Liang, F., Hansen, T.S., Craven, M.B., Bowman, C.L., Renning, C.M.,  
Nierman, W., Fraser, C.M., Martin, G.B., Giovannoni, J.J., and Tanksley  
S.D.  
Generation of ESTs from tomato fruit tissue, breaker stage  
Unpublished (2000)  
Contact: Cuci  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.

FEATURES  
Location/Qualifiers  
source  
1..618  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="CLEG26C9"  
/clone\_lib="tomato breaker fruit, TIGR"  
/tissue\_type="Pericarp"  
/dev\_stage="breaker"  
/lab\_host="SOLR"  
/note="Vector: pBluescriptSKmCvAdapt; Site\_1: EcoRI;  
Site\_2: XhoI; Fruit were harvested at the breaker stage  
(first sign of lycopene accumulation on the blossom end of  
the fruit). Fruit were cut in half and the seeds and  
locules were discarded prior to freezing the pericarp."  
BASE COUNT 196 a 115 c 149 g 157 t 1 others  
ORIGIN  
Query Match 14.6%; Score 546; DB 10; Length 618;  
Best Local Similarity 100.0%; Pred. No. 2.3e-278;  
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2830 CAAAAAATTCGTGAGCTCTTTCAATTGACGTGGACTTCCAAAGACTGGTGTCCCGC 2889  
DB 31 CAAAAAATTCGTGAGCTCTTTCAATTGACGTGGACTTCCAAAGACTGGTGTCCCGC 90  
QY 2890 TGAATACCATCTCATGTCGCCCTTAAGAATACCAGACTTCATGGATAGCCGACAA 2949  
DB 91 TGAATACCATCTCATGTCGCCCTTAAGAATACCAGACTTCATGGATAGCCGACAA 150  
QY 2950 GACCAGCTATATCTCAGAAGAGTATTGGAAGCTTTTCAGGAAAGTGAAGGACAAAGC 3009  
DB 151 GACCAGCTATATCTCAGAAGAGTATTGGAAGCTTTTCAGGAAAGTGAAGGACAAAGC 210  
QY 3010 ACCTCAGGCTAGCTCTATCGGACCTTCACAAGAGATGTTCAAGGAGATCATATATGC 3069  
DB 211 ACCTCAGGCTAGCTCTATCGGACCTTCACAAGAGATGTTCAAGGAGATCATATATGC 270  
QY 3070 TGATATGGAAGTTGATGATTTGAAGATTCATTCAGCAAGCTTTTGACTACAAAATGA 3129  
DB 271 TGATATGGAAGTTGATGATTTGAAGATTCATTCAGCAAGCTTTTGACTACAAAATGA 330  
QY 3130 ATATGACAACAAGCTGGGTAAATTTAATGGACTACTATGSCATAAAAACAGAGGCTGAAAT 3189  
DB 331 ATATGACAACAAGCTGGGTAAATTTAATGGACTACTATGSCATAAAAACAGAGGCTGAAAT 390  
QY 3190 ACTTAGTGGGATGATTAAGAGCATCAAAAACCTTTTACCAGCAAGAGATGCTGAGGC 3249  
DB 391 ACTTAGTGGGATGATTAAGAGCATCAAAAACCTTTTACCAGCAAGAGATGCTGAGGC 450  
QY 3250 CATTAGTGTGCTGTGAGGCGCTTTCAGGAAGGAGGAGAGAGCTGTTTCAAGAGGCGTAA 3309  
DB 451 CATTAGTGTGCTGTGAGGCGCTTTCAGGAAGGAGGAGAGAGAGCTGTTTCAAGAGGCGTAA 510

```

QY 3310 TGATATAGATGACATGTTACCAAGGCTTGGCTTGGTACACGTTTACATATCATCTAC 3369
|||||
Db 511 TGATATAGATGACATGTTACCAAGGCTTGGCTTGGTACACGTTTACATATCATCTAC 570
QY 3370 ATATTG 3375
|||||
Db 571 ATATTG 576

RESULT 3
AW933692
LOCUS EST359535 tomato fruit mature green, TAMU Lycopersicon esculentum
DEFINITION cDNA clone cLEF56C18 5', mRNA sequence.
ACCESSION AW933692
VERSION AW933692.1 GI:8109093
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 525)
AUTHORS Alcalá, J., Vrebalov, J., White, R., Matern, A.L., Holt, I.E., Liang, F.,
Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S., Ronning, C.M.,
Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
TITLE Generation of ESTs from tomato fruit tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
source
1..525
Location/Qualifiers
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEF56C18"
/clone_lib="tomato fruit mature green, TAMU"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLEF - Fruit were tagged at the 1cm stage and
harvested 3-5 days prior to ripening. Fruit were cut in
half to verify the seeds were indeed 'immature' and the
seeds and locules were discarded prior to freezing the
pericarp"
BASE COUNT 151 a 111 c 121 g 142 t
ORIGIN
Query Match 14.1%; Score 525; DB 10; Length 525;
Best Local Similarity 100.0%; Pred. No. 3.3e-267;
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2464 TCCATGCTTCATCCCGGTGATTCGTTTAAAGGCTGTAATGTTCCGAGCGCTGCA 2523
|||||
Db 1 TCCATGCTTCATCCCGGTGATTCGTTTAAAGGCTGTAATGTTCCGAGCGCTGCA 60
QY 2524 CCATGCTGATGTTGTTGTTATTCCTCAGAAAGGAAAAAGACCTCATCCGAATGAATG 2583
|||||
Db 61 CCATGCTGATGTTGTTGTTATTCCTCAGAAAGGAAAAAGACCTCATCCGAATGAATG 120
QY 2584 TTCTGGAGTATTTGGATGGGATATCTACTTTGTTCTGGGATCAAGACATGATCCC 2643
|||||
Db 121 TTCTGGAGTATTTGGATGGGATATCTACTTTGTTCTGGGATCAAGACATGATCCC 180
QY 2644 GCCAAGGCAAGTCCACCGCATGAATATCCTCCAGCACCACGACATAGTTGGACCATGA 2703
|||||
Db 181 GCCAAGGCAAGTCCACCGCATGAATATCCTCCAGCACCACGACATAGTTGGACCATGA 240

```

```

QY 2704 TGTCACAATTGAGGAAGTTGAAGAGTACTTACCACACTATATTGTGAATGACAGTTGGG 2763
|||||
Db 241 TGTCACAATTGAGGAAGTTGAAGAGTACTTACCACACTATATTGTGAATGACAGTTGGG 300
QY 2764 AATCATAGCAAAATGCCCATGTGTATTTGCAGACAGAGAACCTGATATGCCCATGAGTGA 2823
|||||
Db 301 AATCATAGCAAAATGCCCATGTGTATTTGCAGACAGAGAACCTGATATGCCCATGAGTGA 360
QY 2824 TCCATGCAAAAAAAGCTTGTGAGCTCTTTTCAATTGCAGTGGACTTTTCCAAAGACTGGTGT 2883
|||||
Db 361 TCCATGCAAAAAAAGCTTGTGAGCTCTTTTCAATTGCAGTGGACTTTTCCAAAGACTGGTGT 420
QY 2884 TCCCGCTGAAATACCATCTCAGTTGGCCCTTAAAGATAATCCAGACTTCATGGATAAGCC 2943
|||||
Db 421 TCCCGCTGAAATACCATCTCAGTTGGCCCTTAAAGATAATCCAGACTTCATGGATAAGCC 480
QY 2944 GGACAAGACCAAGCTATATCTCAGAAAGAGTATTGGAAGCTTTT 2988
|||||
Db 481 GGACAAGACCAAGCTATATCTCAGAAAGAGTATTGGAAGCTTTT 525

RESULT 4
AW96465
LOCUS EST265896 tomato callus, TAMU Lycopersicon esculentum cDNA clone
DEFINITION cLEF15G23, mRNA sequence.
ACCESSION AW96465
VERSION AW96465.1 GI:5602355
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 421)
AUTHORS Alcalá, J., Vrebalov, J., White, R., Matern, A.L., Vision, T., Holt, I.E.,
Liang, F., Upton, J., Craven, M.B., Bowman, C.L., Ahn, S., Ronning,
C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
TITLE Generation of ESTs from tomato callus tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
source
1..421
Location/Qualifiers
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEF15G23"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni laboratory; cLEF - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato callus EST Library"
BASE COUNT 117 a 85 c 101 g 118 t
ORIGIN
Query Match 11.3%; Score 421; DB 9; Length 421;
Best Local Similarity 100.0%; Pred. No. 6.5e-212;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2464 TCCATGCTTCATCCCGGTGATTCGTTTAAAGGCTGTAATGTTCCGAGCGCTGCA 2523
|||||

```

```

Db 1 TCCATGCTTCATCCTCGTGATATTCCTGTTTAAAGGCTGTAATGTTTCGAGCGCTGCA 60
QY 2524 CCACATGGTAGATGTTGTTGTTATCCCTCAGAAAGGAAAAAGACCTCATCCGAATG 2583
Db 61 CCACATGGTAGATGTTGTTGTTATCCCTCAGAAAGGAAAAAGACCTCATCCGAATG 120
QY 2584 TTCTGGGAGTGATTGGATGGGATATCTACTTTGTTTGGTGATCAAGACATGATCC 2643
Db 121 TTCTGGGAGTGATTGGATGGGATATCTACTTTGTTTGGTGATCAAGACATGATCC 180
QY 2644 GCCAAGCAAGTCAGCCGATGGAATATCCTCCAGCACCAGCAATACATGTCGACATGA 2703
Db 181 GCCAAGCAAGTCAGCCGATGGAATATCCTCCAGCACCAGCAATACATGTCGACATGA 240
QY 2704 TGTCACAATGAGGAAGTGAAGAGTACTTCACTTCTTGTGATGAAAGACAGTTGGG 2763
Db 241 TGTCACAATGAGGAAGTGAAGAGTACTTCACTTCTTGTGATGAAAGACAGTTGGG 300
QY 2764 AATCATGCAAAATGCCCATGCTGATTTGCGAGACAGAGAACCTGATATGCCATGAGTGA 2823
Db 301 AATCATGCAAAATGCCCATGCTGATTTGCGAGACAGAGAACCTGATATGCCATGAGTGA 360
QY 2824 TCCATGCAAAATGCTGAGCTCTTTTCAATTGCGAGTGGACTTCCAAAGACCTGGTGT 2883
Db 361 TCCATGCAAAATGCTGAGCTCTTTTCAATTGCGAGTGGACTTCCAAAGACCTGGTGT 420
QY 2884 T 2884
Db 421 T 421

RESULT 5
LOCUS BE433646 570 bp mRNA linear EST 18-MAY-2001
DEFINITION EST400175 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
clone cLEG16J9, mRNA sequence.
ACCESSION BE433646
VERSION BE433646.1 GI:9431489
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 570)
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Niernan,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
S.D.
TITLE Generation of ESTs from tomato fruit tissue, breaker stage
JOURNAL Unpublished (2000)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
source Location/Qualifiers
1..570
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEG16J9"
/tissue_type="tomato breaker fruit, TIGR"
/dev_stage="breaker"
/lab_host="SOLR"
/note="Vector: pBluescriptSKmCvAdapt; Site_1: EcoRI;
Site_2: XhoI; Fruit were harvested at the breaker stage
(first sign of lycopene accumulation on the blossom end
the fruit). Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."

```

```

BASE COUNT 157 a 112 c 134 g 167 t
ORIGIN
Query Match 10.7%; Score 398; DB 10; Length 570;
Best Local Similarity 100.0%; Pred. NO. 1.2e-139;
Matches 398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2309 TTGATCAATCCAGAACCTTGGATATGCTCAGGTCTTTGTTTCAGTTTACTGCTGGA 2368
Db 173 TTGATCAATCCAGAACCTTGGATATGCTCAGGTCTTTGTTTCAGTTTACTGCTGGA 232
QY 2369 CATGGAGAGTTTCTGACGATTTACATCCATTTAATACAGCAGATCCACCACAGTAAT 2428
Db 233 CATGGAGAGTTTCTGACGATTTACATCCATTTAATACAGCAGATCCACCACAGTAAT 292
QY 2429 TTCAATCTGAAGGAATGTTGTTTGCAGAAATCCATGCTTGCATCCTGCTGATAT 2488
Db 293 TTCAATCTGAAGGAATGTTGTTTGCAGAAATCCATGCTTGCATCCTGCTGATAT 352
QY 2489 CGTGTTTTAAAGGCTGTAAATGTTCCAGGCTGCACACATGTTAGATTTGTTTATTC 2548
Db 353 CGTGTTTTAAAGGCTGTAAATGTTCCAGGCTGCACACATGTTAGATTTGTTTATTC 412
QY 2549 CCTCAGAAAGGAAAGACCTCATCCGAATGATGTTCTGGGAGTGATTTGGATGGGAT 2608
Db 413 CCTCAGAAAGGAAAGACCTCATCCGAATGATGTTCTGGGAGTGATTTGGATGGGAT 472
QY 2609 ATCTACTTTGTTGCTGGGATCAAGACATGATCCGCCAAGCAAGTCCAGCCGATGAA 2668
Db 473 ATCTACTTTGTTGCTGGGATCAAGACATGATCCGCCAAGCAAGTCCAGCCGATGAA 532
QY 2669 TATCCTCCAGCACCAGCATACAGTTGGACCATGATGT 2706
Db 533 TATCCTCCAGCACCAGCATACAGTTGGACCATGATGT 570

RESULT 6
LOCUS AW932918 537 bp mRNA linear EST 18-MAY-2001
DEFINITION EST358761 tomato fruit mature green, TAMU Lycopersicon esculentum
cDNA clone CLEF50H7 5', mRNA sequence.
ACCESSION AW932918
VERSION AW932918.1 GI:8108319
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 537)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M.,
Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.J.
TITLE Generation of ESTs from tomato fruit tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
source Location/Qualifiers
1..537
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEF50H7"
/tissue_type="tomato fruit mature green, TAMU"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:

```

XhoI; cLEF - Fruit were tagged at the 1cm stage and harvested 3-5 days prior to ripening. Fruit were cut in half to verify the seeds were indeed 'immature' and the seeds and locules were discarded prior to freezing the pericarp"

BASE COUNT 177 a 112 c 118 g 130 t

ORIGIN

Query Match 9.6%; Score 358; DB 10; Length 537;  
Best Local Similarity 99.8%; Pred. No. 2.2e-178;  
Matches 408; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2635 CATGATCCGGCAAGGCAAGTCCAGCCGATGGATATATCTCCAGACCCAGCCATACAGTT 2694  
|||||  
DB 1 CATGATCCGGCAAGGCAAGTCCAGCCGATGGATATATCTCCAGACCCAGCCATACAGTT 60  
|||||

QY 2695 GGACCATGATGTCACAAATGAGGAAGTTGAAGAGTACTTCCAACTATATTTGTAATGA 2754  
|||||  
DB 61 GGACCATGATGTCACAAATGAGGAAGTTGAAGAGTACTTCCAACTATATTTGTAATGA 120  
|||||

QY 2755 CAGTTTGGGAATCATAGCAAAATGCCATGTCGTATTTTCGACACAGAGAACCTGATATGCC 2814  
|||||  
DB 121 CAGTTTGGGAATCATAGCAAAATGCCATGTCGTATTTTCGACACAGAGAACCTGATATGCC 180  
|||||

QY 2815 CATGATGATCCATCCAAAAACTTGGCTGAGCTCTTTTCAATTCAGTGGACTTCCAAA 2874  
|||||  
DB 181 CATGATGATCCATCCAAAAACTTGGCTGAGCTCTTTTCAATTCAGTGGACTTCCAAA 240  
|||||

QY 2875 GACTGGTGTCCGCGTGAATACCATCTCAGTTGCGCCCTAAAGAAATACCCAGACTTCAT 2934  
|||||  
DB 241 GACTGGTGTCCGCGTGAATACCATCTCAGTTGCGCCCTAAAGAAATACCCAGACTTCAT 300  
|||||

QY 2935 GGATAGCCGGACAGACAGCATATATCTCAGAAAGAGTTATTTGGAAGCTTTTCAGGAA 2994  
|||||  
DB 301 GGATAGCCGGACAGACAGCATATATCTCAGAAAGAGTTATTTGGAAGCTTTTCAGGAA 360  
|||||

QY 2995 AGTGAAGCAAGACACCTCAGGCTAGCTATCGACCTTCAAGA 3043  
|||||

DB 361 AGTGAAGCAAGACACCTCAGGCTAGCTATCGGACCTTCAAGA 409  
|||||

RESULT 7  
AW649811  
LOCUS  
DEFINITION  
EST328265 tomato germinating seedlings, TAMU Lycopersicon  
esculentum cDNA clone cLEI9L3 5', mRNA sequence.

ACCESSION  
AW649811  
VERSION  
AW649811.1 GI:7411049  
KEYWORDS  
EST.  
SOURCE  
tomato.

ORGANISM  
Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

REFERENCE  
1 (bases 1 to 597)  
Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,  
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,  
Nierman,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and Tanksley  
,S.D.  
Generation of ESTs from germinating tomato seed  
Unpublished (2000)  
Contact: CUGI  
Clemson University  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.

TITLE  
JOURNAL  
COMMENT  
Unpublished (2000)  
Contact: CUGI  
Clemson University  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.

FEATURES  
Location/Qualifiers  
1..597  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"

/clone="cLEI9L3"  
/clone\_lib="tomato germinating seedlings, TAMU"  
/tissue\_type="whole seedlings"  
/dev\_stage="7 days post imbibition"  
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:  
XhoI; 7 days post imbibition on water-agar. Mixed stage  
whole germinating seedlings from seed coat emergence up  
to two centimeters in seeds not showing obvious signs of  
germination were discarded."

BASE COUNT 178 a 122 c 138 g 159 t

ORIGIN

Query Match 9.1%; Score 341; DB 10; Length 597;  
Best Local Similarity 99.7%; Pred. No. 2.4e-169;  
Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2572 TCCGAATGAATGTTCTGGGAGTGTATTTGGATGGGGATATCTACTTTGTTGCTGGGATCA 2631  
|||||  
DB 149 TCCGAATGAATGTTCTGGGAGTGTATTTGGATGGGGATATCTACTTTGTTGCTGGGATCA 208  
|||||

QY 2632 AGACATATCCCGCAAGCAAGTCCAGCCGATGGATATCTCCAGACCCAGCATACA 2691  
|||||  
DB 209 AGACATATCCCGCAAGCAAGTCCAGCCGATGGATATCTCCAGACCCAGCATACA 268  
|||||

QY 2692 GTTGGACCATGATGTCTACAATTGAGGAAGTTGAAGAGTACTTCCACCACTATATTGTGAA 2751  
|||||  
DB 269 GTTGGACCATGATGTCTACAATTGAGGAAGTTGAAGAGTACTTCCACCACTATATTGTGAA 328  
|||||

QY 2752 TGACAGTTTGGGAATCATAGCAAAATGCCATGTCGTATTTGCGAGACAGAGAACTGATAT 2811  
|||||  
DB 329 TGACAGTTTGGGAATCATAGCAAAATGCCATGTCGTATTTGCGAGACAGAGAACTGATAT 388  
|||||

QY 2812 GGCATCAGTGTCCCTGCAAAAACCTTGCTGAGCTCTTTTCAATTGCACTGACATTTCC 2871  
|||||  
DB 389 GGCATCAGTGTCCCTGCAAAAACCTTGCTGAGCTCTTTTCAATTGCACTGACATTTCC 448  
|||||

QY 2872 AAAGACTGGTGTTCGCCCTGAAATACCATCTCAGTTGCGCCCTAAAGAAATACCCAGCTT 2931  
|||||  
DB 449 AAAGACTGGTGTTCGCCCTGAAATACCATCTCAGTTGCGCCCTAAAGAAATACCCAGCTT 508  
|||||

QY 2932 CATGGATAAGCCGGACAGACCAAGCTATATCT 2963  
|||||

DB 509 CATGGATAAGCCGGACAGACCAAGCTATATCT 540  
|||||

RESULT 8  
BF112386  
LOCUS  
DEFINITION  
EST439976 tomato breaker fruit Lycopersicon esculentum cDNA clone  
cLEG41011 5' sequence, mRNA sequence.

ACCESSION  
BF112386  
VERSION  
BF112386.1 GI:10942076  
KEYWORDS  
EST.  
SOURCE  
tomato.

ORGANISM  
Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

REFERENCE  
1 (bases 1 to 367)  
Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,  
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,  
Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley  
,S.D.  
Generation of ESTs from tomato fruit tissue, breaker stage  
Unpublished (2000)  
Contact: CUGI  
Clemson University  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
Location/Qualifiers  
1..367

FEATURES  
source

/organism="Lycopersicon esculentum"  
 /cultivar="TA496"  
 /db\_xref="taxon:4081"  
 /clone="cLEG41011"  
 /clone\_lib="tomato breaker fruit"  
 /tissue\_type="pericarp"  
 /dev\_stage="breaker"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescriptSKmCudapt; Site\_1: EcoRI;  
 Site\_2: XhoI; supplier: Boyce Thompson Institute;  
 sequencing: The Institute for Genomic Research. Fruit  
 were harvested at the breaker stage (first sign of  
 lycopene accumulation on the blossom end of fruit). Fruit  
 were cut in half and the seeds and locules were discarded  
 prior to freezing the pericarp."  
 102 a 75 c 89 g 101 t

BASE COUNT 102 a 75 c 89 g 101 t  
 ORIGIN  
 Query Match 8.5%; Score 316; DB 12; Length 367;  
 Best Local Similarity 99.7%; Pred. No. 4.4e-156; Indels 0; Gaps 0;  
 Matches 366; Conservative 0; Mismatches 1;

Qy 2464 TCCATGCTTGCATCCTGGTGATATTCGTTTAAAGGCTGTAATGTTTCGAGCGTGCA 2523  
 Db 1 TCCATGCTTGCATCCTGGTGATATTCGTTTAAAGGCTGTAATGTTTCGAGCGTGCA 60  
 Qy 2524 CCATGGTAGATTGTTGTAATTCCTCAGAAAGGAAAGACCCTCATCCGATGATG 2583  
 Db 61 CCATGGTAGATTGTTGTAATTCCTCAGAAAGGAAAGACCCTCATCCGATGATG 120  
 Qy 2584 TTCGGGAGTAGTTGGATGGGATATCTACTTTTGGTGGATCAAGACATGATCCC 2643  
 Db 121 TTCGGGAGTAGTTGGATGGGATATCTACTTTTGGTGGATCAAGACATGATCCC 180  
 Qy 2644 GCCAAGCAATCCAGCGATGGAATATCTCCAGCAGCCAGCATACAGTTGGACATGA 2703  
 Db 181 GCCAAGCAATCCAGCGATGGAATATCTCCAGCAGCCAGCATACAGTTGGACATGA 240  
 Qy 2704 TGTCACAATTGAGGAAGTTGAAGAGTACTTCCACCACTATATTTGTAATGACAGTTGGG 2763  
 Db 241 TGTCACAATTGAGGAAGTTGAAGAGTACTTCCACCACTATATTTGTAATGACAGTTGGG 300  
 Qy 2764 AATCATAGCAATGCCATGTCGTATTTGCAGACAGAGACCTGATATGGCCATGATGA 2823  
 Db 301 AATCATAGCAATGCCATGTCGTATTTGCAGACAGAGACCTGATATGGCCATGATGA 360  
 Qy 2824 TCCATGC 2830  
 Db 361 TCCATGC 367

RESULT 9  
 BG594667  
 LOCUS BG594667 725 bp mRNA linear EST 12-APR-2001  
 DEFINITION EST493345 cSTS Solanum tuberosum cDNA clone cST88119 5' sequence,  
 mRNA sequence.  
 ACCESSION BG594667  
 VERSION BG594667.1 GI:13612807  
 KEYWORDS EST.  
 SOURCE potato.  
 ORGANISM Solanum tuberosum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 1 (bases 1 to 725)  
 van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chiemingo,A.,  
 Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.  
 Generations of ESTs from sprouting potato eyes  
 Unpublished (2000)  
 TITLE Contact: Cathy Ronning  
 JOURNAL The Institute for Genomic Research  
 COMMENT For clone info: please contact Research Genetics, Libraries  
 Division tel 1-800-711-6195, email cdna@resgen.com

Seq primer: M13F-R.  
 Location/Qualifiers  
 1. 725  
 /organism="Solanum tuberosum"  
 /cultivar="Kennebec"  
 /db\_xref="taxon:4113"  
 /clone="cSTS8119"  
 /clone\_lib="cSTS"  
 /tissue\_type="sprouting eyes from tubers"  
 /dev\_stage="12-14 weeks post harvest"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
 XhoI; Various sizes of sprouting eyes (2mm to 15mm) were  
 taken from tubers. The tubers were incubated at 26C in the  
 dark for 2-3 weeks prior to sprouting. The eyes were  
 frozen in liquid nitrogen immediately upon removal from  
 tubers."

BASE COUNT 228 a 140 c 170 g 187 t  
 ORIGIN

Query Match 5.2%; Score 195; DB 12; Length 725;  
 Best Local Similarity 99.6%; Pred. No. 1.1e-91;  
 Matches 245; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 2699 CATGATGTCCACCAATTGAGGAAGTTGAAGAGTACTTCCACCACTATATTTGAATGACAGT 2758  
 Db 212 CATGATGTCCACCAATTGAGGAAGTTGAAGAGTACTTCCACCACTATATTTGAATGACAGT 271  
 Qy 2759 TTGGGAATCATAGCAATGCCATGTCGTTTGGAGACAGAGAACCTGATATGCCATG 2818  
 Db 272 TTGGGAATCATAGCAATGCCATGTCGTTTGGAGACAGAGAACCTGATATGCCATG 331  
 Qy 2819 AGTGATCCCATGCAAAAACCTTGTGAGCTCTTTTCAATTGCAAGTGGACTTCCAAAGACT 2878  
 Db 332 AGTGATCCCATGCAAAAACCTTGTGAGCTCTTTTCAATTGCAAGTGGACTTCCAAAGACT 391  
 Qy 2879 GGTGTTCCCGCTGAATACCATCTCGATTCGCGCCCTAAAGATATCCAGACTTCATGGAT 2938  
 Db 392 GGTGTTCCCGCTGAATACCATCTCGATTCGCGCCCTAAAGATATCCAGACTTCATGGAT 451  
 Qy 2939 AAGCCG 2944  
 Db 452 AAGCCG 457

RESULT 10  
 BG111608  
 LOCUS BG111608 521 bp mRNA linear EST 22-JUL-2002  
 DEFINITION EST597184 mixed potato tissues Solanum tuberosum cDNA clone STMCB87  
 5' end mRNA sequence.  
 ACCESSION BG111608  
 VERSION BG111608.2 GI:21915035  
 KEYWORDS EST.  
 SOURCE potato.  
 ORGANISM Solanum tuberosum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 1 (bases 1 to 521)  
 Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,  
 Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and  
 Karamycheva,S.A.  
 Generation of a set of potato cDNA clones for microarray analyses  
 Unpublished (2002)  
 On Apr 17, 2002 this sequence version replaced gi:20163570.  
 TITLE Contact: Robin Buell  
 JOURNAL The Institute for Genomic Research  
 COMMENT 9712 Medical Center Dr, Rockville, MD 20850, USA  
 Email: potato@tigr.org  
 This clone is available through the Research Genetics, contact the  
 Research Genetics for further information 1-800-711-6195 or  
 cdna@resgen.com  
 Seq primer: T3.

```

FEATURES
source
1. .521
Location/Qualifiers
/organism="Solanum tuberosum"
/cultivar="Kennebec or Binjete"
/db_xref="taxon:4113"
/clone_lib="STMCB87"
/tissue_type="mixed potato tissues"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Combination of untreated and Phytophthora
infestans-treated libraries of stolons, leaves, leaflets,
axillary buds of stem explants, petioles, germinating eyes
, tubers, or roots."
BASE COUNT 155 a 106 c 121 g 139 t
ORIGIN
Query Match 3.8%; Score 141; DB 14; Length 521;
Best Local Similarity 99.5%; Pred. No. 5.4e-63;
Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2753 GACAGTTGGGAATCATAGCAAAATGCCATGTCGTATTTGCAGACAGAGAACCTGATATG 2812
Db 274 GACAGTTGGGAATCATAGCAAAATGCCATGTCGTATTTGCAGACAGAGAACCTGATG 333
QY 2813 GCCATGAGTGATCCATGCAAAAACCTGCTGAGCTCTTTCAATTGCGAGTGACTTTCCA 2872
Db 334 GCCATGAGTGATCCATGCAAAAACCTGCTGAGCTCTTTCAATTGCGAGTGACTTTCCA 393
QY 2873 AAGACTGGTGTCCCGCTGAAATACCAATCAGTTCAGTGGCCCTAAAGAAATACCCAGACTTC 2932
Db 394 AAGACTGGTGTCTGCTGAAATACCAATCAGTTCAGTGGCCCTAAAGAAATACCCAGACTTC 453
QY 2933 ATGGATAAGCGG 2944
Db 454 ATGGATAAGCGG 465

RESULT 11
BE920043
LOCUS BE920043 610 bp mRNA linear EST 02-OCT-2000
DEFINITION BE920043 610 bp mRNA linear EST 02-OCT-2000
ACCESSION BE920043
VERSION BE920043.1 GI:10446119
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
1 (bases 1 to 610)
van der Hoeven,R.S., Bezzerides,J., Holt,I.E., Liang,F., Cho,J.,
Uterback,T., Hansen,C.L., Doan,B., Bougri,O., Buell,C.R., Ronning
,C.M., Fry,W.E., Tanksley,S.D. and Baker,B.
Generation of ESTs from potato leaves and petioles
Unpublished (2000)
Contact: Cathy Ronning
The Institute for Genomic Research
For clone request: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cda@resgen.com.
Location/Qualifiers
1. .610
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone_lib="ESTB3D19"
/tissue_type="leaflets and petioles"
/dev_stage="8 weeks old plants"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Tissue was supplied by Dr. Fry (Cornell University)."
FEATURES
source
1. .610
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone_lib="ESTB3D19"
/tissue_type="leaflets and petioles"
/dev_stage="8 weeks old plants"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Tissue was supplied by Dr. Fry (Cornell University)."
FEATURES
source
1. .434
/organism="Solanum tuberosum"
/cultivar="Kennebec or Binjete"
/db_xref="taxon:4113"
/clone_lib="STMCB87"
/tissue_type="mixed potato tissues"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Combination of untreated and Phytophthora
infestans-treated libraries of stolons, leaves, leaflets,
axillary buds of stem explants, petioles, germinating eyes
, tubers, or roots."
BASE COUNT 129 a 95 c 73 g 137 t
ORIGIN
Query Match 3.2%; Score 121; DB 14; Length 434;

```

Leaflets and petioles were isolated from 8 week old greenhouse grown plants. The plants were watered and fertilized freely. The tissue was immediately frozen in liquid nitrogen."

BASE COUNT 181 a 121 c 146 g 162 t

ORIGIN

Query Match 3.8%; Score 141; DB 12; Length 610;

Best Local Similarity 99.5%; Pred. No. 5.5e-63;

Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2753 GACAGTTGGGAATCATAGCAAAATGCCATGTCGTATTTGCAGACAGAGAACCTGATATG 2812

Db 274 GACAGTTGGGAATCATAGCAAAATGCCATGTCGTATTTGCAGACAGAGAACCTGATATG 333

QY 2813 GCCATGAGTGATCCATGCAAAAACCTGCTGAGCTCTTTCAATTGCGAGTGACTTTCCA 2872

Db 334 GCCATGAGTGATCCATGCAAAAACCTGCTGAGCTCTTTCAATTGCGAGTGACTTTCCA 393

QY 2873 AAGACTGGTGTCCCGCTGAAATACCAATCAGTTCAGTTCAGTGGCCCTAAAGAAATACCCAGACTTC 2932

Db 394 AAGACTGGTGTCTGCTGAAATACCAATCAGTTCAGTTCAGTGGCCCTAAAGAAATACCCAGACTTC 453

QY 2933 ATGGATAAGCGG 2944

Db 454 ATGGATAAGCGG 465

RESULT 12

BE920043

LOCUS BE920043 434 bp mRNA linear EST 17-APR-2002

DEFINITION BE920043 434 bp mRNA linear EST 17-APR-2002

ACCESSION BE920043

VERSION BE920043.1 GI:20163571

KEYWORDS EST.

SOURCE potato.

ORGANISM Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 434)

Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C., Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and Karamycheva,S.A.

Generation of a set of potato cDNA clones for microarray analyses

Unpublished (2002)

Contact: Robin Buell

The Institute for Genomic Research

9712 Medical Center Dr, Rockville, MD 20850, USA

Email: potato@tigr.org

This clone is available through the Research Genetics, contact the Research Genetics for further information 1-800-711-6195 or cda@resgen.com

Seq primer: T7.

Location/Qualifiers

1. .434

/organism="Solanum tuberosum"

/cultivar="Kennebec or Binjete"

/db\_xref="taxon:4113"

/clone\_lib="STMCB87"

/tissue\_type="mixed potato tissues"

/lab\_host="SOLR"

/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; supplier: Combination of untreated and Phytophthora infestans-treated libraries of stolons, leaves, leaflets, axillary buds of stem explants, petioles, germinating eyes , tubers, or roots."

BASE COUNT 129 a 95 c 73 g 137 t

ORIGIN

Query Match 3.2%; Score 121; DB 14; Length 434;

Best Local Similarity 99.4%; Pred. No. 2.3e-52;  
Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3291 CCTGGTTCAGAGCGCTAATGATATAGATGACATGTTACCAAGCGCTTCGGCTGGTACC 3350  
|||||  
Db 419 CTGGGTTCAAGAGCGCTAATGATATAGATGACATGTTACCAAGCGCTTCGGCTGGTACC 360  
|||||  
QY 3351 ACCTTACATATCATCTACATATTTGGGTTGCTCAATCAGGCGTTGAAAAGAGCTCAT 3410  
|||||  
Db 359 ACCTTACATATCATCTACATATTTGGGTTGCTCAATCAGGCGTTGAAAAGAGATCAT 300  
|||||  
QY 3411 TCATTAGCTTCCCTGGTGTGTTAAGACCGACTAATCCAGATTAAGAAGGA 3462  
|||||  
Db 299 TCATTAGCTTCCCTGGTGTGTTAAGACCGACTAATCCAGATTAAGAAGGA 248  
|||||

RESULT 13  
BF052042  
LOCUS  
DEFINITION  
EST437289 tomato developing/immature green fruit Lycopersicon  
esculentum cDNA clone cLEM25110 5' sequence, mRNA sequence.  
ACCESSION  
BF052042  
VERSION  
BF052042.1 GI:10805938  
KEYWORDS  
EST.  
SOURCE  
tomato.  
ORGANISM  
Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.

REFERENCE  
AUTHORS  
Alcala, J., Vrebalov, J., White, R., van der Hoeven, R.S., Holt, I.E.,  
Liang, F., Hansen, T.S., Craven, M.B., Bowman, C.L., Ronning, C.M.,  
Nierman, W., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley  
S.D.

TITLE  
JOURNAL  
COMMENT  
Generation of ESTs from tomato fruit tissue, immature green  
Unpublished (2000)  
Contact: CUGI  
Clemson University  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>.

FEATURES  
Source  
1..403  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLEM25110"  
/clone\_lib="tomato developing/immature green fruit"  
/tissue\_type="fruit"  
/dev\_stage="immature green (5-35 days post-anthesis)"  
/lab\_host="SOLR"  
/note="Vector: pBluescriptSKmCudapt; Site\_1: EcoRI;  
Site\_2: XhoI; Fruit were tagged at 5 dpa (0.5 cm) and  
harvested at 7 day intervals through 35 dpa. Equal masses  
of tissue from each stage were combined (including seeds  
and locules) prior to mRNA isolation."

BASE COUNT  
128 a 68 c 97 g 110 t  
ORIGIN  
Query Match 2.9%; Score 110; DB 12; Length 403;  
Best Local Similarity 100.0%; Pred. No. 1.6e-46;  
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3281 GAGGCAAGAGCGTGGTTCAAGAGCGGTAATGATATAGATGACATGTTACCAAGGCTTCG 3340  
|||||  
Db 222 GAGGCAAGAGCGTGGTTCAAGAGCGGTAATGATATAGATGACATGTTACCAAGGCTTCG 281  
|||||  
QY 3341 GCTTGGTACCAGGTACATATATCATCTACATATTTGGGTTGCTACAATCA 3390  
|||||  
Db 282 GCTTGGTACCAGGTACATATATCATCTACATATTTGGGTTGCTACAATCA 331  
|||||

RESULT 14  
BG597875  
LOCUS

DEFINITION  
EST496553 cSTS Solanum tuberosum cDNA clone cSTS19C12 5' sequence,  
mRNA sequence.  
ACCESSION  
BG597875  
VERSION  
BG597875.1 GI:13616015  
KEYWORDS  
EST.  
SOURCE  
potato.  
ORGANISM  
Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
REFERENCE  
1 (bases 1 to 682)  
AUTHORS  
van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Chieningo, A.,  
Bougr, O., Buell, C.R., Ronning, C., Tanksley, S. and Baker, B.

TITLE  
JOURNAL  
COMMENT  
Generations of ESTs from sprouting potato eyes  
Unpublished (2000)  
Contact: Cathy Ronning  
The Institute for Genomic Research  
For clone info: please contact Research Genetics, Libraries  
Division tel 1-800-711-6195, email [cdna@resgen.com](mailto:cdna@resgen.com)  
Seq primer: M13F-R.

FEATURES  
Location/Qualifiers  
1..682  
Source

/organism="Solanum tuberosum"  
/cultivar="Kennebec"  
/db\_xref="taxon:4113"  
/clone="cSTS19C12"  
/clone\_lib="cSTS"  
/tissue\_type="sprouting eyes from tubers"  
/dev\_stage="12-14 weeks post harvest"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; Various sizes of sprouting eyes (2mm to 15mm) were  
taken from tubers. The tubers were incubated at 26C in the  
dark for 2-3 weeks prior to sprouting. The eyes were  
frozen in liquid nitrogen immediately upon removal from  
tubers."

BASE COUNT  
185 a 108 c 154 g 235 t  
ORIGIN

Query Match 2.7%; Score 101; DB 12; Length 682;  
Best Local Similarity 100.0%; Pred. No. 1.1e-41;  
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 614 TATAAATTCAGCTTTCATATGAAATATATGCGAGGTTGTGCTCCATCGTCATATGGT 673  
|||||  
Db 540 TATAAATTCAGCTTTCATATGAAATATATGCGAGGTTGTGCTCCATCGTCATATGGT 599  
|||||  
QY 674 CAAATGCTCAGTTCTCCTCATACAGTTATTTGGTGCTCC 714  
|||||  
Db 600 CAAATGCTCAGTTCTCCTCATACAGTTATTTGGTGCTCC 640  
|||||

RESULT 15  
BQ507474  
LOCUS

DEFINITION  
EST614889 Generation of a set of potato cDNA clones for microarray  
analyses mixed potato tissues Solanum tuberosum cDNA clone STMG059  
5' end, mRNA sequence.  
ACCESSION  
BQ507474  
VERSION  
BQ507474.2 GI:21923304  
KEYWORDS  
EST.  
SOURCE  
potato.  
ORGANISM  
Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
REFERENCE  
1 (bases 1 to 773)  
AUTHORS  
Buell, C.R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C.,  
Restrepo, S., Griffiths, H., van der Hoeven, R., Tsai, J. and  
Karamycheva, S.A.

BASE COUNT  
773 a 108 c 154 g 235 t  
ORIGIN

Query Match 2.7%; Score 101; DB 12; Length 682;  
Best Local Similarity 100.0%; Pred. No. 1.1e-41;  
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 614 TATAAATTCAGCTTTCATATGAAATATATGCGAGGTTGTGCTCCATCGTCATATGGT 673  
|||||  
Db 540 TATAAATTCAGCTTTCATATGAAATATATGCGAGGTTGTGCTCCATCGTCATATGGT 599  
|||||  
QY 674 CAAATGCTCAGTTCTCCTCATACAGTTATTTGGTGCTCC 714  
|||||  
Db 600 CAAATGCTCAGTTCTCCTCATACAGTTATTTGGTGCTCC 640  
|||||

TITLE  
JOURNAL  
COMMENT  
Generation of a set of potato cDNA clones for microarray analyses  
On Jun 10, 2002 this sequence version replaced gi:21366343.  
Contact: Robin Buell  
The Institute for Genomic Research  
9712 Medical Center Dr, Rockville, MD 20850, USA  
Email: potato@tigr.org  
This clone is available through the Research Genetics, contact the  
Research Genetics for further information 1-800-711-6195 or  
cdna@resgen.com  
Seq primer: T3

FEATURES  
source  
1..773  
/organism="Solanum tuberosum"  
/cultivar="Kennebec Or Binite"  
/db\_xref="taxon:4113"  
/clone="STM059"  
/clone.lib="Generation of a set of potato cDNA clones for  
microarray analyses mixed potato tissues"  
/tissue\_type="mixed tissues"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; supplier: Combination of untreated and Phytophthora  
infestans-treated libraries of stolons, leaves, leaflets,  
axillary buds of stem explants, petioles, germinating eyes  
tubers, or roots."  
BASE COUNT 205 a 125 c 181 g 262 t  
ORIGIN

Query Match 2.6%; Score 98; DB 14; Length 773;  
Best Local Similarity 100.0%; Pred. No. 4.2e-40;  
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 644 TGGCAGGTGCTCCATCGTCATATGGTCAAAATGCTCAGTTTCCTCATACAGTTA 703  
|||||  
Db 577 TGGCAGGTGCTCCATCGTCATATGGTCAAAATGCTCAGTTTCCTCATACAGTTA 636  
QY 704 TTGTGTCTCTCGGATCTATAAGAGACTTGAAGAACTC 741  
|||||  
Db 637 TTGTGTCTCTCGGATCTATAAGAGACTTGAAGAACTC 674

RESULT 16  
AW623969  
LOCUS  
DEFINITION  
EST321914 tomato flower buds 3-8 mm, Cornell University  
Lycopersicon esculentum cDNA clone cTOB13P9 5', mRNA sequence.  
ACCESSION  
AW623969  
VERSION  
AW623969.1 GI:7336996  
KEYWORDS  
EST.  
SOURCE  
tomato.  
ORGANISM  
Lycopersicon esculentum

REFERENCE  
AUTHORS  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 443)  
van der Hoeven, R.S., Bezzerides, J.L., Matern, A.L., Holt, I.E., Liang  
, F., Hansen, T., Craven, M.B., Bowman, C.L., Ronning, C.M., Nierman, W.,  
Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley, S.D.  
Generation of ESTs from tomato flower tissue, 3-8 mm buds  
Unpublished (1999)  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.

FEATURES  
source  
1..443  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"

/clone="cTOB13P9"  
/clone.lib="tomato flower buds 3-8 mm, Cornell University"  
/tissue\_type="flower"  
/dev\_stage="3-8mm buds"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; supplier: Tanksley; Flower buds and flowers were  
taken from greenhouse plants (4-8 wks old, TA496). They  
were immediately frozen in liquid nitrogen and then  
size-separated while remaining frozen."  
BASE COUNT 135 a 101 c 98 g 109 t  
ORIGIN

Query Match 2.5%; Score 95; DB 10; Length 443;  
Best Local Similarity 99.3%; Pred. No. 1.5e-38;  
Matches 145; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2955 GCTATATCTCAGAAAGAGTTATTGGAAGCTTTTCAGGAAAGTGAAGCAACAGCACCTC 3014  
|||||  
Db 293 GCTATATCTCAGAAAGAGTTATTGGAAGCTTTTCAGGAAAGTGAAGCAACAGCACCTC 352  
QY 3015 AGGCTAGCTCTATCGGACCTTCACAGAGATGTTGCAAGGAGATCATATGATGCTGATA 3074  
|||||  
Db 353 ATGCTAGCTCTATCGGACCTTCACAGAGATGTTGCAAGGAGATCATATGATGCTGATA 412  
QY 3075 TGAAGTTGATGGATTTGAAGATTAC 3100  
|||||  
Db 413 TGAAGTTGATGGATTTGAAGATTAC 438

RESULT 17  
BG592782  
LOCUS  
DEFINITION  
EST491460 cSTS Solanum tuberosum cDNA clone cSTS2K4 5' sequence,  
mRNA sequence.  
ACCESSION  
BG592782  
VERSION  
BG592782.1 GI:13610922  
KEYWORDS  
EST.  
SOURCE  
potato.  
ORGANISM  
Solanum tuberosum

REFERENCE  
AUTHORS  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
1 (bases 1 to 529)  
van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Chienmingo, A.,  
Bougr, O., Buell, C.R., Ronning, C., Tanksley, S. and Baker, B.  
Generations of ESTs from sprouting potato eyes  
Unpublished (2000)  
Contact: Cathy Ronning  
The Institute for Genomic Research  
For clone info: please contact Research Genetics, Libraries  
Division tel 1-800-711-6195, email [cdna@resgen.com](mailto:cdna@resgen.com)  
Seq primer: M13F-R.

FEATURES  
source  
1..529  
/organism="Solanum tuberosum"  
/cultivar="Kennebec"  
/db\_xref="taxon:4113"  
/clone.lib="cSTS"  
/clone="cSTS2K4"  
/tissue\_type="sprouting eyes from tubers"  
/dev\_stage="12-14 weeks post harvest"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; Various sizes of sprouting eyes (2mm to 15mm) were  
taken from tubers. The tubers were incubated at 26C in the  
dark for 2-3 weeks prior to sprouting. The eyes were  
frozen in liquid nitrogen immediately upon removal from  
tubers."  
BASE COUNT 168 a 103 c 120 g 138 t  
ORIGIN

Query Match 2.3%; Score 84; DB 12; Length 529;  
Best Local Similarity 99.3%; Pred. No. 1.1e-32;



```

Matches 134; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2715 AGGAAGTTGAAGTACTTCCACCAACTATATTGTGAATGACAGTTGGGAATCATAGCAA 2774
Db 25 AGGAAGTTGAAGTACTTCCACCAACTATATTGTGAATGACAGTTGGGAATCATATCAA 84
QY 2775 ATGCCCATGCGTATTTGCGAGACAGAACCTGATATGGCCATGAGTGATCCATGCAAAA 2834
Db 85 ATGCCCATGCGTATTTGCGAGACAGAACCTGATATGGCCATGAGTGATCCATGCAAAA 144
QY 2835 AACTTGCTGAGCTCT 2849
Db 145 AACTTGCTGAGCTCT 159

RESULT 18
BG889665
LOCUS EST515516 cSTD Solanum tuberosum cDNA clone cSTD15C7 5' sequence,
DEFINITION mRNA sequence.
ACCESSION BG889665
VERSION BG889665
KEYWORDS EST.
SOURCE BG889665.1 GI:14266751
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 393)
AUTHORS van der Hoeven,R., Bezzerides,J., Ewing,E., Cho,J., Chieningo,A.,
Bougr,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
TITLE Generations of ESTs from dormant potato tubers
JOURNAL Unpublished (2001)
COMMENT Contact: Cathy Ronning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13F-R.
FEATURES
source
1..393
Location/Qualifiers
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cSTD15C7"
/tissue_type="dormant tuber"
/dev_stage="one month post-harvest"
/lab_host="SOLR"
/notes="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; This library targets genes expressed in dormant
tubers. This library was made from sections of dormant
tuber, avoiding the buds and epidermis. Tubers were stored
for one month post-harvest at 40C. The tuber was peeled,
well away from the surface. Then it was chopped into 1-2
mm cubes and immediately frozen in liquid nitrogen. This
library is noted as P4 in Tanksley lab notebooks."
BASE COUNT 114 a 60 c 96 g 123 t
ORIGIN
Query Match 2.2%; Score 81; DB 12; Length 393;
Best Local Similarity 100.0%; Pred. No. 4.2e-31;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 270 ATACAGGATGGAACTGATATGTCATGGAGGTTAAACAGTCCAAAGGAGGATCTAGAG 329
Db 152 ATACAGGATGGAACTGATATGTCATGGAGGTTAAACAGTCCAAAGGAGGATCTAGAG 211
QY 330 CATTTGCCAAAGTTCAATTG 350
Db 212 CATTTGCCAAAGTTCAATTG 232
RESULT 19

```

---

```

BE340733 321 bp mRNA linear EST 17-JUL-2000
LOCUS EST344811 potato stolon, Cornell University Solanum tuberosum cDNA
DEFINITION clone cSTAl2M12 similar to RNA-directed RNA polymerase
(Lycopersicon esculentum) PIR[T30819|T30819 RNA-direc. mRNA
sequence.
ACCESSION BE340733
VERSION BE340733.1 GI:9250264
KEYWORDS EST.
SOURCE BE340733.1
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 321)
AUTHORS van der Hoeven,R., Bezzerides,J., Bachem,C., Horvath,B., Visser,R.,
Holt,I.E., Liang,F., Hansen,T.S., Utterback,T., Bowman,C.L., Doan
B., Bougr,O., Buell,C.R., Ronning,C.M., Tanksley,S.D. and Baker
,B.
TITLE Generation of ESTs from potato swelling stolons
JOURNAL Unpublished (1999)
COMMENT Contact: Research Genetics, Libraries Division
Tel: 1-800-711-6195
Email: cdna@resgen.com
5 prime sequence.
FEATURES
source
1..321
Location/Qualifiers
/organism="Solanum tuberosum"
/cultivar="Bintje"
/db_xref="taxon:4113"
/clone="cSTAl2M12"
/clone_lib="potato stolon, Cornell University"
/tissue_type="axillary buds of stem explants, swelling
stolons"
/dev_stage="1 to 3 days"
/lab_host="SOLR"
/notes="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; RNA was supplied by Christian Bachem & Beatrix
Horvath(Laboratory of Plant Breeding, Dept. of Plant
Sciences, Wageningen University, The Netherlands). Total
RNA was isolated from developing axillary buds of potato
nodal stem cuttings cultured on medium for the
introduction of tuber formation as described in Bachem et
al. (Plant Journal 1996). Tissue samples were taken of
stages corresponding to growing stolons and the early
stages of tuber formation."
BASE COUNT 87 a 51 c 80 g 103 t
ORIGIN
Query Match 2.1%; Score 78; DB 10; Length 321;
Best Local Similarity 99.2%; Pred. No. 1.6e-29;
Matches 128; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2243 AAGTTGCTCGATTGGCGACTAGATCAAGAATATTTATTCCAAATGGAAGACAATGATG 2302
Db 172 AAGTTGCTCGATTGGCGACTAGATCAAGAATATTTATTCCAAATGGAAGACAATGATG 231
QY 2303 GGATGTTGGATGAATCCAGAACCTTGGATATGGTCAGGTGTTTTCAGTTACTGCT 2362
Db 232 GGATGTTGGATGAATCCAGAACCTTGGATATGGTCAGGTGTTTTCAGTTACTGCT 291
QY 2363 GCTGGACAT 2371
Db 292 GCTGGACAT 300

RESULT 20
AJ487433
LOCUS AJ487433 Solanum tuberosum cv. Provita Solanum tuberosum cDNA clone
DEFINITION p3f8, mRNA sequence.
ACCESSION AJ487433
VERSION AJ487433.1 GI:22022186

```

KEYWORDS EST.  
 SOURCE potato.  
 ORGANISM Solanum tuberosum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 REFERENCE 1 (bases 1 to 550)  
 AUTHORS Gebhardt, C., Walkemeier, B., Henselewski, H., Barakat, A., Delseny, M. and Stueber, K.  
 TITLE Comparative mapping between potato (Solanum tuberosum) and Arabidopsis thaliana reveals structurally conserved domains and ancient duplications in the potato genome  
 JOURNAL unpublished (2002)  
 COMMENT Contact: Gebhardt C  
 MPI for Breeding Research  
 Carl-von-Linne Weg 10, Cologne, D-50829, Germany.  
 FEATURES source  
 1. .550  
 /organism="Solanum tuberosum"  
 /cultivar="Provita"  
 /db\_xref="taxon:4113"  
 /map="v"  
 /clone="P3f8"  
 /clone\_lib="Solanum tuberosum cv. Provita"  
 BASE COUNT 182 a 91 c 125 g 152 t  
 ORIGIN  
 Query Match 1.9%; Score 71; DB 9; Length 550;  
 Best Local Similarity 99.2%; Pred. No. 9.2e-26;  
 Matches 121; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 3341 GCTTGGTACCGGTACATATCATCTACATATTGGGGTGTCTACATCAGGGTTGAAA 3400  
 |||||  
 Db 222 GCTTGGTACCGGTACATATCATCTACATATTGGGGTGTCTACATCAGGGTTGAAA 281  
 QY 3401 AGAGTCATTCATTAGCTTCCCTGGTGTGTATGACACGCTAATCCAGATTAAAGAG 3460  
 |||||  
 Db 282 AGAGTCATTCATTAGCTTCCCTGGTGTGTATGACACGCTAATCCAGATTAAAGAG 341  
 QY 3461 GA 3462  
 ||  
 Db 342 GA 343  
 RESULT 21  
 BQ509665/c  
 LOCUS  
 DEFINITION EST617080 Generation of a set of potato cDNA clones for microarray analyses mixed potato tissues Solanum tuberosum cDNA clone STWH37 3', end, mRNA sequence.  
 ACCESSION BQ509665  
 VERSION BQ509665.2 GI:21925399  
 KEYWORDS EST.  
 SOURCE potato.  
 ORGANISM Solanum tuberosum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 REFERENCE 1 (bases 1 to 340)  
 AUTHORS Buell, C.R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C., Restrepo, S., Griffiths, H., van der Hoeven, R., Tsai, J. and Karamycheva, S.A.  
 TITLE Generation of a set of potato cDNA clones for microarray analyses  
 JOURNAL unpublished (2002)  
 COMMENT On Jun 10, 2002 this sequence version replaced gi:21368534.  
 Other\_ESTs: EST617079  
 Contact: Robin Buell  
 The Institute for Genomic Research  
 9712 Medical Center Dr, Rockville, MD 20850, USA  
 Email: potato@tigr.org  
 This clone is available through the Research Genetics, contact the Research Genetics for further information 1-800-711-6195 or

cdna@resgen.com  
 Seq primer: T7  
 FEATURES source  
 1. .340  
 /organism="Solanum tuberosum"  
 /cultivar="Kennebec or Binjfe"  
 /db\_xref="taxon:4113"  
 /clone="STWHE37"  
 /clone\_lib="Generation of a set of potato cDNA clones for microarray analyses mixed potato tissues"  
 /tissue\_type="mixed tissues"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; supplier: Combination of untreated and Phytophthora infestans-treated libraries of stolons, leaves, leaflets, axillary buds of stem explants, petioles, germinating eyes, tubers, or roots."  
 BASE COUNT 107 a 53 c 55 g 115 t  
 ORIGIN  
 Query Match 1.5%; Score 57; DB 14; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-18;  
 Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3406 TCATTTGATTAGCTTCCCTGGTGTGTATGACACGCTAATCCAGATTAAAGGA 3462  
 |||||  
 Db 315 TCATTTGATTAGCTTCCCTGGTGTGTATGACACGCTAATCCAGATTAAAGGA 259  
 RESULT 22  
 BQ827639  
 LOCUS  
 DEFINITION BQ827639.1 541 bp mRNA linear EST 07-AUG-2002  
 9d58f03.y1 Moss EST library PPAS Physcomitrella patens cDNA clone  
 PEP\_SOURCE\_ID:PPAS041205 5' similar to TR:Q9SG02 Q9SG02 PUTATIVE  
 RNA-DIRECTED RNA POLYMERASE. ;, mRNA sequence.  
 ACCESSION BQ827639  
 VERSION BQ827639.1 GI:221333265  
 KEYWORDS EST.  
 SOURCE Physcomitrella patens.  
 ORGANISM Physcomitrella patens  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.  
 REFERENCE 1 (bases 1 to 541)  
 AUTHORS Quatrano, R., Bashardes, S., Cove, D., Cuming, A., Knight, C., Clifton, S., Marra, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Underwood, K., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Steptoe, M., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R., Waterston, R. and Wilson, R.  
 TITLE Leeds/Wash U Moss EST Project  
 JOURNAL unpublished (1999)  
 COMMENT Contact: Ralph Quatrano  
 Leeds/Wash U Moss EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 Libraries were constructed by Dr. Stavros Bashardes as part of the Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and Washington Univ. in St. Louis (USA) DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 447.  
 FEATURES source  
 1. .541  
 /organism="Physcomitrella patens"  
 /db\_xref="taxon:3218"  
 /clone="PEP\_SOURCE\_ID:PPAS041205"  
 /clone\_lib="Moss EST library PPAS"  
 /dev\_stage="protonemata, 7day old ABA-treated"  
 /lab\_host="E.coli DH10b"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:

ECORI; 7-day-old pronemal tissue was incubated 10-4M ABA for 16 hours before RNA isolation. cDNA was synthesised from Amersham's cDNA Synthesis plus kit and ligated with EcorI-NotI linkers for cloning in the EcorI site of Lambda ZapII (Stratagene). After packaging, the library was propagated in E. coli XL-I Blue cells and amplified. The library was excised by mass excision in SOLR cells and ampicillin resistant transformants selected. Approximately 1,000,000 colonies were grown and recovered. The double stranded plasmid library was recovered and used to transform DH10b cells by electroporation. Clones corresponding to abundant transcripts were identified by colony hybridization using a cDNA probe derived from untreated pronemal tissue, and eliminated from the library, by rearraying."

BASE COUNT 133 a 113 c 141 g 154 t  
 ORIGIN  
 Query Match 0.6%; Score 24; DB 14; Length 541;  
 Best Local Similarity 100.0%; Pred. No. 0.91;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2458 AAAAAATCCATGCTTCATCCTGG 2481  
 |||||  
 Db 349 AAAAAATCCATGCTTCATCCTGG 372

RESULT 23  
 BH258389  
 LOCUS  
 DEFINITION CH230-112K20-TV CHORI-230 Segment 1 Rattus norvegicus genomic clone  
 CH230-112K20, DNA sequence.  
 ACCESSION BH258389  
 VERSION BH258389.1 GI:17160712  
 KEYWORDS GSS.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE 1 (bases 1 to 348)  
 AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn  
 A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de  
 Jong,P. and Fraser,C.M.  
 TITLE Rat BAC End Sequences from Library CHORI-230 EcorI segment  
 JOURNAL Unpublished (1999)  
 COMMENT Other GSSs: CH230-112K20.TJ  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the rat BAC library CHORI-230  
 (http://www.chori.org/bacpac/rat230.htm). For BAC library  
 availability, please contact Pieter de Jong (pjejong@mail.cho.org).  
 Clones may be purchased from BACPAC Resources  
 (http://www.chori.org/bacpac/orering\_information.htm). BAC end  
 page: http://www.tigr.org/tdb/bac\_ends/rat/bac\_end\_intro.html  
 Plate: 112 row: K column: 20  
 Seq primer: T7  
 Class: BAC ends.  
 Location/Qualifiers  
 1. .348  
 /organism="Rattus norvegicus"  
 /strain="BN/SSNHsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone="CH230-112K20"  
 /clone\_lib="CHORI-230 Segment 1"  
 /sex="Female"  
 /cell\_type="Brain"  
 /note="Vector: pTARBAC2.1; Site\_1: EcorI; Site\_2: EcorI;

FEATURES  
 source

1. .348  
 /organism="Rattus norvegicus"  
 /strain="BN/SSNHsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone="CH230-112K20"  
 /clone\_lib="CHORI-230 Segment 1"  
 /sex="Female"  
 /cell\_type="Brain"  
 /note="Vector: pTARBAC2.1; Site\_1: EcorI; Site\_2: EcorI;

CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by  
 Pieter de Jong"  
 BASE COUNT 90 a 63 c 52 g 143 t  
 ORIGIN  
 Query Match 0.6%; Score 23; DB 17; Length 348;  
 Best Local Similarity 100.0%; Pred. No. 2.9;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 572 TTGAAGAAATTTTCTCTTTT 594  
 |||||  
 Db 239 TTGAAGAAATTTTCTCTTTT 261

RESULT 24  
 BQ488250  
 LOCUS  
 DEFINITION 39-E8847-006-003-M09-T3 Sugar beet MP1Z-ADIS-006 Lambda Zap II  
 library Beta vulgaris cDNA clone M-9-3, mRNA sequence.  
 ACCESSION BQ488250  
 VERSION BQ488250.1 GI:21332870  
 KEYWORDS EST.  
 SOURCE Beta vulgaris.  
 ORGANISM Beta vulgaris  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Caryophyllidae; Caryophyllales; Chenopodiaceae; Beta.

REFERENCE 1 (bases 1 to 571)  
 AUTHORS Bellin,D., Werber,M., Theis,T., Weisshaar,B. and Schneider,K.  
 TITLE EST sequencing, annotation and macroarray expression analysis of  
 more than 3000 sugar beet cDNAs identifies genes with root-specific  
 expression pattern  
 JOURNAL Unpublished (2002)  
 COMMENT ADIS DNA core facility at MP1Z  
 Contact: Weisshaar B  
 Max-Planck-Institute for Plant Breeding Research  
 Carl-von-Linne Weg 10, 50829 Koeln, Germany  
 Fax: 00492215062851  
 Email: weisshaar@mpiz-koeln.mpg.de  
 Seq primer: T3 'AATTACCCCTCACTAAAGG'  
 High quality sequence stop: 571.  
 Location/Qualifiers  
 1. .571  
 /organism="Beta vulgaris"  
 /db\_xref="taxon:161934"  
 /clone="M-9-3"  
 /clone\_lib="Sugar beet MP1Z-ADIS-006 Lambda Zap II  
 library"  
 /dev\_stage="4 week old pot-grown plants"  
 /note="Organ: shoot and root; Vector: pBluescript SK- from  
 lambda Zap II; cDNA (lambda Zap-II) library from sugar  
 beet, whole plant mRNA, prepared using the Stratagene  
 UniZAP cDNA kit, cloning sites EcorI-XhoI, primer sites  
 and orientation:  
 rev-T3-Saci-SK-EcorI-GGCACGAGG-5pr-cDNA-polyA-XhoI-KpnI-T7  
 -uni"

FEATURES  
 source  
 1. .571  
 /organism="Beta vulgaris"  
 /db\_xref="taxon:161934"  
 /clone="M-9-3"  
 /clone\_lib="Sugar beet MP1Z-ADIS-006 Lambda Zap II  
 library"  
 /dev\_stage="4 week old pot-grown plants"  
 /note="Organ: shoot and root; Vector: pBluescript SK- from  
 lambda Zap II; cDNA (lambda Zap-II) library from sugar  
 beet, whole plant mRNA, prepared using the Stratagene  
 UniZAP cDNA kit, cloning sites EcorI-XhoI, primer sites  
 and orientation:  
 rev-T3-Saci-SK-EcorI-GGCACGAGG-5pr-cDNA-polyA-XhoI-KpnI-T7  
 -uni"

BASE COUNT 172 a 103 c 128 g 168 t

Query Match 0.6%; Score 23; DB 14; Length 571;  
 Best Local Similarity 100.0%; Pred. No. 3.1;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2207 TTCTTTCAATGATGTGCAAAAC 2229  
 |||||  
 Db 544 TTCTTTCAATGATGTGCAAAAC 566

RESULT 25  
 BB013867/c  
 LOCUS  
 DEFINITION BB013867 RIKEN full-length enriched, adult male testis (DH10B) Mus  
 musculus cDNA clone 4930472H2 3', mRNA sequence.  
 265 bp mRNA linear EST 22-JUN-2000

```

ACCESSION      BS013867
VERSION        BS013867.1
KEYWORDS       GI:8137026
SOURCE         house mouse.
ORGANISM       Mus musculus

REFERENCE
AUTHORS        Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
               P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
               Hirozane,T., Hori,F., Ishi,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
               Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
               Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,
               Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
               Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,
               Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sobabe,Y., Sugahara,Y.,
               Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomioka,N., Toya
               T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,
               Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
               M., Muramatsu,M. and Hayashizaki,Y.
TITLE          RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL        Unpublished (2000)
COMMENT        Contact: Yoshihide Hayashizaki
               Laboratory for Genome Exploration Research Group, RIKEN Genomic
               Sciences Center(GSC), Yokohama Institute
               The Institute of Physical and Chemical Research (RIKEN)
               1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
               Tel: 81-45-503-9222
               Fax: 81-45-503-9216
               URL: http://genome.gsc.riken.go.jp/
               Email: genome-gsc.riken.go.jp,
               Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
               N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
               Thermostabilization and thermoactivation of thermostable enzymes by
               trehalose and its application for the synthesis of full length
               cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
               Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
               Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
               Y. and Hayashizaki,Y.
               Automated filtration-based high-throughput plasmid preparation
               system. Genome Res. 9 (5), 463-470 (1999)
               Carninci,P. and Hayashizaki,Y.
               High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
               19-44 (1999)
               Please visit our web site (http://genome.rtc.riken.go.jp) for
               further details.

FEATURES       Location/Qualifiers
               source
               1..265
               /organism="Mus musculus"
               /strain="C57BL/6J"
               /db_xref="taxon:10090"
               /clone="4930472H22"
               /clone_lib="RIKEN full-length enriched, adult male testis
               (DH10B)"
               /sex="male"
               /tissue_type="testis"
               /dev_stage="adult"
               /lab_host="DH10B"
               /note="Site.1: Sali; Site.2: BamHI; cDNA library was
               prepared and sequenced in Mouse Genome Encyclopedia
               Project of Genome Exploration Research Group in Riken
               Genomic Sciences Center and Genome Science Laboratory in
               RIKEN. Division of Experimental Animal Research in Riken
               contributed to prepare mouse tissues. 1st strand cDNA was
               primed with a primer [5'
               GAGAGAGAAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was
               prepared by using trehalose thermo-activated reverse
               transcriptase and subsequently enriched for full-length by
               cap-trapper. Second strand cDNA was prepared with the
               primer adapter of sequence [5'
               GAGAGAGAATTCGAGTTAATTAATTAATCCGCCCCCCCCC 3']. cDNA
               was cloned into the XhoI and BamHI sites. Vector: a
               modified pBluescript KS(+) after bulk excision from Lambda

```

```

BASE COUNT     91 a      37 c      39 g      98 t
ORIGIN
Query Match      0.6%      Score 22; DB 10; Length 265;
Best Local Similarity 100.0%; Pred. No. 9.4; 0; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 574 GAAGAAATTTTCTTTCTTTTCTTTT 595
      |||||||
Db 150 GAAGAAATTTTCTTTCTTTTCTTTT 129

RESULT 26
AV045785/c
LOCUS          AV045785
DEFINITION     AV045785 Mus musculus adult C57BL/6J testis Mus musculus cDNA clone
               1700048K22, mRNA sequence.
ACCESSION      AV045785
VERSION        AV045785.2
KEYWORDS       EST.
SOURCE         house mouse.
ORGANISM       Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
               1 (bases 1 to 308)
AUTHORS        Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
               Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara
               A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
               Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsumu,H., Oda,H., Owa,C.,
               Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sobabe,Y., Sugahara
               Y., Suzuki,H., Suzuki,H., Tateo,M., Tomaru,Y., Tomioka,N.,
               Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,
               Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
TITLE          RIKEN Mouse ESTs
JOURNAL        Unpublished (1999)
COMMENT        Contact: Chie Owa
               Genome Science Laboratory
               RIKEN
               3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
               Tel: 81-298-36-9145
               Fax: 81-298-36-9098
               Email: genome-res@rtc.riken.go.jp
               Thermostabilization and thermoactivation of thermostable enzymes by
               trehalose and its application for the synthesis of full length cDNA
               (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
               Transcriptional sequencing: A method for DNA sequencing using RNA
               polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
               Please visit our web site (http://genome.rtc.riken.go.jp) for
               further details.
               Location/Qualifiers
               1..308
               /organism="Mus musculus"
               /strain="C57BL/6J"
               /db_xref="taxon:10090"
               /clone="1700048K22"
               /clone_lib="Mus musculus adult C57BL/6J testis"
               /sex="male"
               /tissue_type="testis"
               /dev_stage="adult"
               /note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia
               ) with a modified polylinker; Site.1: Not I; Site.2: Eco
               RI; 1st strand cDNA was primed with a Not I - oligo(dT)
               primer [5'
               TGTACCAATCTGAAGTGGGAGCGCGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTTTT
               T 3']; double-stranded cDNA was ligated to Eco RI
               adaptors (Pharmacia), digested with Not I and cloned into
               the Not I and Eco RI sites of the modified pT73 vector.
               RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
               constructed and normalized by Bento Soares and M. Fatima
               Bonaldo."
               107 a      44 c      52 g      105 t
BASE COUNT
ORIGIN

```

Query Match 0.6%; Score 22; DB 9; Length 308;  
 Best Local Similarity 100.0%; Pred. No. 9.6;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 574 GAAGAAATTTTCTTTT 595  
 Db 188 GAAGAAATTTTCTTTT 167

RESULT 27  
 AV045898/c 312 bp mRNA linear EST 23-NOV-1999  
 LOCUS  
 DEFINITION AV045898 Mus musculus adult C57BL/6J testis Mus musculus cDNA clone  
 1700049G07, mRNA sequence.  
 ACCESSION AV045898  
 VERSION AV045898.2 GI:4865563  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 312)  
 AUTHORS Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,  
 Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara  
 A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,  
 Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C.,  
 Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara  
 Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomimaga, N.,  
 Watanabe, S., Yeganeh, M., Yamamura, T., Yokota, T., Yoshino, M.,  
 Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.

TITLE RIKEN Mouse ESTs  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Chie Owa  
 Genome Science Laboratory  
 RIKEN  
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
 Tel: 81-298-36-9145  
 Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp  
 Thermostabilization and thermostabilization of thermolabile enzymes by  
 trehalose and its application for the synthesis of full length cDNA  
 (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
 Transcriptional sequencing: A method for DNA sequencing using RNA  
 polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
 further details.

FEATURES Location/Qualifiers  
 1..312  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="1700049G07"  
 /clone\_lib="Mus musculus adult C57BL/6J testis"  
 /sex="male"  
 /tissue\_type="testis"  
 /dev\_stage="adult"  
 /note="Organ: mammary gland; Vector: p7T3D-Pac (Pharmacia  
 ) with a modified polylinker; Site\_1: Not I; Site\_2: Eco  
 RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
 primer [5',  
 TGTACCAATCTGAAGTGGGAGCGCGGAATGTTTCTTTTCTTTTCTTTTCTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified p7T3 vector.  
 RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
 constructed and normalized by Bento Soares and M. Fatima  
 Bonaldo."  
 BASE COUNT 109 a 46 c 53 g 104 t

Query Match 0.6%; Score 22; DB 9; Length 312;  
 Best Local Similarity 100.0%; Pred. No. 9.6;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 574 GAAGAAATTTTCTTTT 595  
 Db 192 GAAGAAATTTTCTTTT 171

RESULT 28  
 AV210762/c 332 bp mRNA linear EST 30-OCT-1999  
 LOCUS  
 DEFINITION AV210762 RIKEN full-length enriched, adult male testis Mus musculus  
 cDNA clone 1700126B05 3', mRNA sequence.

ACCESSION AV210762  
 VERSION AV210762.1 GI:6151249  
 KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 332)

AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,  
 Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F.,  
 Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai  
 C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Kai  
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,  
 Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata  
 Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H.,  
 Suzuki, H., Takahashi, F., Tateno, M., Tomimaga, N., Tsunoda, Y.,  
 Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T.,  
 Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Konno, H., et al. 1999)  
 Unpublished (1999)  
 CONTACT: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.go.jp/  
 URL: <http://genome.gsc.riken.go.jp/>  
 Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y.,  
 Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki  
 Y., Transcriptional sequencing: A method for DNA sequencing using RNA  
 polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
 Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki  
 Y. and Hayashizaki, Y.  
 Automated filtration-based high-throughput plasmid preparation  
 system. Genome Res. 9 (5), 463-470 (1999)  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
 19-44 (1999)  
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
 further details.

TITLE RIKEN Mouse ESTs  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.go.jp/  
 URL: <http://genome.gsc.riken.go.jp/>  
 Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y.,  
 Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki  
 Y., Transcriptional sequencing: A method for DNA sequencing using RNA  
 polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
 Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki  
 Y. and Hayashizaki, Y.  
 Automated filtration-based high-throughput plasmid preparation  
 system. Genome Res. 9 (5), 463-470 (1999)  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
 19-44 (1999)  
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
 further details.

FEATURES Location/Qualifiers  
 1..332  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="1700126B05"  
 /clone\_lib="RIKEN full-length enriched, adult male testis"  
 /sex="male"  
 /tissue\_type="testis"  
 /dev\_stage="adult"  
 /lab\_host="SOLR"  
 /note="Site\_1: XhoI; Site\_2: BamHI; cDNA library was  
 prepared and sequenced in Mouse Genome Encyclopedia  
 Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in  
 RIKEN. Division of Experimental Animal Research in Riken  
 contributed to prepare mouse tissues. 1st strand cDNA was



Soares, M.B. 'Normalization and subtraction: two approaches to facilitate gene discovery', Genome Research 6:791-806, 1996. The first strand synthesis used a NotI-dt18 primer; double stranded cDNAs were ligated to EcoRI adapters, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pTZ19-Pac vector. The library contained approximately  $7.2 \times 10^5$  recombinants, with average insert sizes of 1-1.5 kb."

BASE COUNT 175 a 87 c 207 t

ORIGIN  
Query Match 0.6%; Score 22; DB 10; Length 580;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1049 CTTGGCCAGCATTAATGTC 1070

DB 489 CTTGGCCAGCATTAATGTC 468

RESULT 31  
AV332932  
LOCUS  
DEFINITION  
AV332932 RIKEN full-length enriched, adult male medulla oblongata  
Mus musculus cDNA clone 6330543D05 3' similar to X64146  
G domesticus mRNA BT61, mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
EST.  
AV332932.1 GI:6372984  
house mouse.

REFERENCE  
AUTHORS  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 275)  
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,  
Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F.,  
Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai,  
C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,  
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,  
Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata,  
Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H.,  
Suzuki, H., Takahashi, F., Tateno, M., Tomimaga, N., Tsunoda, Y.,  
Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T.,  
Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.  
RIKEN Mouse ESs (Konno, H., et al. 1999)  
Unpublished (1999)

TITLE  
JOURNAL  
COMMENT  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y.,  
Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y., and Hayashizaki,  
Y.

Transcriptional sequencing: A method for DNA sequencing using RNA  
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,  
Y., and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
13-44 (1999)

please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

FEATURES  
Location/Qualifiers  
1..275  
/organism="Mus musculus"

/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="6330543D06"  
/clone\_lib="RIKEN full-length enriched, adult male medulla  
oblongata"  
/sex="male"  
/tissue\_type="medulla oblongata"  
/dev\_stage="adult"  
/lab\_host="DH10B"

/note="Site 1: SalI; Site 2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN, Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3']. cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 10.0 and subtraction to Rot = 100.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGATTCGAGTTAAATTAATTCCTCCCTCCCTCC  
3']. cDNA was cloned into the XhoI and BamHI sites.  
Vector: a modified pBluescript KS(+) after bulk excision  
from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:  
BamHI"

BASE COUNT 60 a 43 c 46 g 126 t

ORIGIN  
Query Match 0.6%; Score 21; DB 10; Length 275;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 576 AGAAATTTTCTTTCTTTTAT 596

DB 227 AGAAATTTTCTTTCTTTAT 247

RESULT 32

AV207607/c

LOCUS

DEFINITION

AV207607 RIKEN full-length enriched, adult male testis muscle

CDNA clone 1700096A08 3', mRNA sequence.

ACCESSION

AV207607

VERSION

AV207607.1 GI:6148460

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 282)

Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,

Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F.,

Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai,

C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,

Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,

Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata,

Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H.,

Suzuki, H., Takahashi, F., Tateno, M., Tomimaga, N., Tsunoda, Y.,

Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T.,

Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.

RIKEN Mouse ESs (Konno, H., et al. 1999)

Unpublished (1999)

TITLE

JOURNAL

COMMENT

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsuura, S., Carninci, P., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.  
 Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
 Itoh, M., Kitsumai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.  
 Automated filtration-based high-throughput plasmid preparation System. Genome Res. 9 (5), 463-470 (1999)  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

#### FEATURES

##### Location/Qualifiers

```
1..282
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="1700096A08"
/clone_lib="RIKEN full-length enriched, adult male testis"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="SOLR"
/note="Site 1: XhoI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAAGATCCAGAGCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAAGCGCGCAATTAATTCGAGTTAATTAATCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites."
```

99 a 40 c 45 g 98 t

#### BASE COUNT

##### ORIGIN

```
Query Match 0.6%; Score 21; DB 9; Length 282;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 575 AAGAAATTTTTCCTTTTA 595
|||||
Db 161 AAGAAATTTTTCCTTTTA 141
```

#### RESULT 33

##### LOCUS

```
BG379160/c 284 bp mRNA linear EST 12-MAR-2001
DEFINITION
UI-R-BT1-bnk-e-09-0-UI.s1 UI-R-BT1 Rattus norvegicus cDNA clone
```

UI-R-BT1-bnk-e-09-0-UI 3', mRNA sequence.

##### ACCESSION

##### VERSION

##### KEYWORDS

##### SOURCE

##### ORGANISM

Rattus norvegicus  
 Norway rat.  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

#### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### MEDLINE

##### COMMENT

Genome Res. 6 (9), 791-806 (1996)  
 9704477  
 Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping

University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: mssoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics ([www.resgen.com](http://www.resgen.com)) Seq primer: M13 Forward  
 POLYA=Yes.

#### FEATURES

##### Source

```
1..284
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BT1-bnk-e-09-0-UI"
/clone_lib="UI-R-BT1"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The library UI-R-BT1 is a subtracted library derived from a mixture of the following tissues: hippocampus, thalamus, mid-brain, medulla, corpus striatum, cerebral cortex and testis. For a detailed description of the library from which this clone was derived, please visit our web site at ratseq.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)  

TAG_SEQ=None found"
```

BASE COUNT 78 a 63 c 58 g 85 t

#### ORIGIN

```
Query Match 0.6%; Score 21; DB 12; Length 284;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 2770 AGCAATGCCATGTCGTATT 2790

|||||

Db 103 AGCAATGCCATGTCGTATT 83

#### RESULT 34

##### LOCUS

##### DEFINITION

```
BB446629 323 bp mRNA linear EST 19-JUL-2000
BB446629 RIKEN full-length enriched, 9 days embryo Mus musculus
cDNA clone D030056M06.3, similar to U39066 Murine MAP kinase kinase
```

6c mRNA, mRNA sequence.

##### ACCESSION

##### VERSION

##### KEYWORDS

##### SOURCE

##### ORGANISM

Mus musculus  
 house mouse.  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

#### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### COMMENT

1 (bases 1 to 323)  
 Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Oho, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suganara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Konno, H., et al.)  
 Unpublished (2000)  
 Contact: Yoshihide Hayashizaki



Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Sueniro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Thermotabilization and thermoactivation of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kitsuina, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

# FEATURES

## source

```
1. 323
   Location/Qualifiers
     /organism="Mus musculus"
     /db_xref="taxon:10090"
     /clone="D030056M06"
     /clone_lib="RIKEN full-length enriched, 9 days embryo"
     /dev_stage="9 days embryo"
     /lab_host="DH10B"
     /note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCTCGAGTTAATTAATATCCGCCGCCGCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified p Bluescript KS(+) after bulk excision from Lambda FLC I."
```

# BASE COUNT

```
81 a 75 c 57 g 110 t
```

Query Match 0.6%; Score 21; DB 10; Length 323;

Best Local Similarity 100.0%; Pred. No. 33; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1293 TTTATTTCGTGTCAGAGG 1313

```
|||||
```

Db 56 TTTATTTCGTGTCAGAGG 76

RESULT 35

AV202914

LOCUS AV202914 Yuji Kohara unpublished cDNA Caenorhabditis elegans cdna clone yk556el2 5', mRNA sequence. EST 26-JUL-1999

DEFINITION AV202914.1 GI:5586685

ACCESSION AV202914

VERSION AV202914.1

KEYWORDS EST.

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea

; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 381)

AUTHORS Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Mitsuki, H., Nishigaki, A., Motohashi, T., Zeng, Q., Watanabe, H., Sugimoto, A., Sano

M., Miyata, A., Mitani, Y., Iida, K., Uesugi, H., Sugiyama, Y. and Nomoto, H.  
Expressed genes in C.elegans  
Unpublished (1999)  
Contact: Yuji Kohara  
Genome Biology Lab.  
National Institute of Genetics  
Yata 1111, Mishima, Shizuoka 411, Japan  
Tel: 81-559-81-6854  
Fax: 81-559-81-6855  
Email: ykoha@elab.nig.ac.jp.  
Location/Qualifiers

# FEATURES

## source

```
1. 381
   /organism="Caenorhabditis elegans"
   /strain="CB1489 him-8(e1489)"
   /db_xref="taxon:6239"
   /clone="yk556el2"
   /clone_lib="Yuji Kohara unpublished cDNA"
   /sex="hermaphrodite, male"
   /tissue_type="whole animal"
   /dev_stage="varied"
```

BASE COUNT 118 a 66 c 77 g 119 t 1 others  
ORIGIN

Query Match 0.6%; Score 21; DB 9; Length 381;

Best Local Similarity 100.0%; Pred. No. 34; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 TCATATGAAATATATGCGAG 649

```
|||||
```

Db 33 TCATATGAAATATATGCGAG 53

RESULT 36

BE581464/c

LOCUS BE581464.1

DEFINITION K446b12.y1 TEN95TM-SSR Strongyloides stercoralis cDNA 5', mRNA sequence. 426 bp mRNA linear EST 09-MAY-2001

ACCESSION BE581464

VERSION BE581464.1

KEYWORDS EST.

SOURCE Strongyloides stercoralis.

ORGANISM Strongyloides stercoralis

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;

Panagrolaimoidea; Strongyloidea; Strongyloidea.

REFERENCE 1 (bases 1 to 426)

AUTHORS McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, V., Runko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

THE Washington Univ. Nematode EST Project, 1999

Unpublished (1999)

Contact: McCarter JP

The Washington Univ. Nematode EST Project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

The library was constructed by Dr. Thomas Nutman and colleagues of

NIH (tnutman@nih.gov). DNA sequencing by: Washington

University Genome Sequencing Center St. Louis.

High quality sequence stop: 411.

Location/Qualifiers

1. 426

/organism="Strongyloides stercoralis"

/strain="Rhabditiform larvae obtained from gerbils"

/db\_xref="taxon:6248"

/clone\_lib="TEN95TM-SSR"

/lab\_host="XL-1 Blue MRF" (Stratagene)"

/note="Vector: Lambda Uni-ZAP XR (Stratagene); Site\_1: EcoRI; Site\_2: XhoI; mRNA was purified from 2 x 10E3 rhaditiform larvae which had been isolated from gerbills experimentally infected with larvae originally isolated from experimentally infected dogs. cDNA was constructed and using adaptors, was cloned unidirectionally into the vector from the EcoRI site to the XhoI site. The library has an unamplified titer of 1 x 10E5 pfu/ml and an amplified, undiluted titer of 9 x 10E11 pfu/ml. The average insert size of the unamplified library is 675 bp (range, 100-1700)."

BASE COUNT 168 a 47 c 76 g 135 t  
ORIGIN

Query Match 0.6%; Score 21; DB 10; Length 426;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1509 ATAAATAATTGAATTCCTG 1529  
|||||  
DB 250 ATAAATAATTGAATTCCTG 270  
|||||

RESULT 37  
LOCUS A0498746  
DEFINITION HS\_5075\_B2\_G09\_SP6E RPCI-11 Human Male BAC Library Homo sapiens  
ACCESSION A0498746  
VERSION A0498746.1 GI:4698858  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 440)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380389  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu  
Plate: 651 row: N column: 18  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 440.  
Location/Qualifiers  
1..440  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate=651 Col=18 Row=N"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"  
BASE COUNT 159 a 88 c 101 g 90 t 2 others

ORIGIN

Query Match 0.6%; Score 21; DB 17; Length 440;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2788 ATTTCAGACAGAGAACCTGA 2808  
|||||  
DB 386 ATTTCAGACAGAGAACCTGA 406  
|||||

RESULT 38  
LOCUS BG705217/c  
DEFINITION 602688174F1 NIH\_MGC\_95 Homo sapiens cDNA clone IMAGE:4820322 5', mRNA sequence.  
ACCESSION BG705217  
VERSION BG705217.1 GI:13979333  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 484)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@remail.nih.gov  
Tissue procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA library preparation: Michael J. Brownstein (NHGRI), Shiraki Toshituki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Plate: LLAM10725 row: d column: 19  
High quality sequence stop: 398.  
Location/Qualifiers  
1..484  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4820322"  
/clone\_lib="NIH\_MGC\_95"  
/tissue.type="hippocampus"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the cap-trapper method (Carninci, in preparation). Library constructed by W. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH-MGC Library."  
BASE COUNT 140 a 76 c 115 g 153 t  
ORIGIN

Query Match 0.6%; Score 21; DB 12; Length 484;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 574 GAAGAAATTTTCTTTCTTTT 594  
|||||  
DB 476 GAAGAAATTTTCTTTCTTTT 456  
|||||

RESULT 39  
LOCUS BI693834  
DEFINITION 603342457F1 NC1\_CGAP\_Mam2 Mus musculus cDNA clone IMAGE:5370135 5', mRNA sequence.  
BASE COUNT 500 bp mRNA linear EST 18-SEP-2001

ACCESSION BI693834  
 VERSION BI693834.1 GI:15656450  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 500)  
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps-femail.nih.gov](mailto:cgaps-femail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: LAM11942 row: a column: 16  
 High quality sequence start: 5  
 High quality sequence stop: 500.  
 FEATURES source  
 LOCATION/Qualifiers  
 1..500  
 /organism="Mus musculus"  
 /strain="FVB/N-3"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5370135"  
 /clone\_lib="NCI\_CGAP\_Mam2"  
 /tissue\_type="tumor, biopsy sample"  
 /dev\_stage="5 months"  
 /lab\_host="Dh108"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"  
 BASE COUNT 165 a 101 c 103 g 131 t  
 ORIGIN  
 Query Match 0.6%; Score 21; DB 13; Length 500;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2694 TGGACCATGATGTCACAAATG 2714  
 Db 237 TGGACCATGATGTCACAAATG 217  
 RESULT 40  
 AZ930706  
 LOCUS 504 bp DNA linear GSS 01-APR-2001  
 DEFINITION 474.dhz57g01.s1 Saccharomyces unisporus NRRL Y-1556 Saccharomyces unisporus genomic clone 474.dhz57g01.s1, DNA sequence.  
 ACCESSION AZ930706  
 VERSION AZ930706.1 GI:13501615  
 KEYWORDS GSS.  
 SOURCE Saccharomyces unisporus.  
 ORGANISM Saccharomyces unisporus  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 REFERENCE 1 (bases 1 to 504)  
 AUTHORS Clifton, P.F., Hillier, L.W., Fulton, L., Graves, T., Miner, T., Gish, W.R., Waterston, R.H. and Johnston, M.  
 TITLE Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Johnston M  
 Department of Genetics  
 Washington University Medical School  
 Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA  
 Tel: 314 362 2735  
 Fax: 314 362 7855

Email: [mjgenetics.wustli.edu](mailto:mjgenetics.wustli.edu)  
 Class: random plasmid subclone.  
 Location/Qualifiers  
 1..504  
 /organism="Saccharomyces unisporus"  
 /strain="NRRL Y-1556 (CBS 398)"  
 /db\_xref="taxon:27294"  
 /clone="474.dhz57g01.s1"  
 /clone\_lib="Saccharomyces unisporus NRRL Y-1556"  
 /note="Random genomic sequence"  
 BASE COUNT 174 a 71 c 75 g 184 t  
 ORIGIN  
 Query Match 0.6%; Score 21; DB 17; Length 504;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2270 AGAATATTTATTCCTCAATGGA 2290  
 Db 172 AGAATATTTATTCCTCAATGGA 192  
 RESULT 41  
 BI038396/c  
 LOCUS 521 bp mRNA linear EST 14-JUN-2001  
 DEFINITION RCS-NT0266-180101-021-A06 NTO266 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BI038396  
 VERSION BI038396.1 GI:14445022  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 521)  
 AUTHORS Dias Neto, E., Garcia Correa, R., Vertovski-Almeida, S., Briones, M.R., Nazai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., Delivelira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.  
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5&t2=RC5-NT0266-180101-021-A06&t3=2001-01-18&t4=1>)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 483.  
 Location/Qualifiers  
 1..521  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NT0266"  
 /dev\_stage="Adult"  
 /note="Organ: nervous\_tumor; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (J.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
 BASE COUNT 163 a 89 c 148 g 121 t  
 ORIGIN

```

Query Match          0.6%; Score 21; DB 13; Length 521;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 579 AATTTTTCCTTTTATCTA 599
      |||
Db 215 AATTTTTCCTTTTATCTA 195

RESULT 42
BI041199
LOCUS
DEFINITION RC5-NT0266-220201-013-E01 NT0266 Homo sapiens cDNA, mRNA sequence.
ACCESSION BI041199
VERSION BI041199.1 GI:14447825
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 529)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL PROC.Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=RC5&t2=RC5-NT0266-
220201-013-E01&t3=2001-02-22&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 528.
Location/Qualifiers
1..529
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0266"
/dev_stage="Adult"
/note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 106 a 143 c 86 g 194 t
ORIGIN
Query Match          0.6%; Score 21; DB 13; Length 529;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 579 AATTTTTCCTTTTATCTA 599
      |||
Db 306 AATTTTTCCTTTTATCTA 326

RESULT 43
AQ985038/c

```

```

LOCUS
DEFINITION AQ985038 630 bp DNA linear GSS 30-JAN-2000
PCI-23-32267.TJ PCI-23 Mus musculus genomic clone PCI-23-32267,
DNA sequence.
ACCESSION AQ985038
VERSION AQ985038.1 GI:6818243
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 630)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library PCI-23
JOURNAL Unpublished (1999)
COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library PCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 322 row: G column: 7
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..630
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="PCI-23-32267"
/clone_lib="PCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 179 a 129 c 123 g 199 t
ORIGIN
Query Match          0.6%; Score 21; DB 17; Length 630;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2915 AAAGAAATACCCAGACTTCATG 2935
      |||
Db 428 AAAGAAATACCCAGACTTCATG 408

RESULT 44
AZ866201/c
LOCUS
DEFINITION AZ866201 637 bp DNA linear GSS 21-FEB-2001
2M0176B12R Mouse 10Kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0176B12 R, DNA sequence.
ACCESSION AZ866201
VERSION AZ866201.1 GI:13067272
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 637)

```

**AUTHORS**  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmood, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
and Wright, D., Weiss, R.  
**TITLE**  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
**JOURNAL**  
Unpublished (2000)  
**COMMENT**  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0176 row: B column: 12  
Seq primer: CACACAGGACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 637.  
**FEATURES**  
source  
Location/Qualifiers  
1. .637  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UGC2M0176B12"  
/clone\_lib="Mouse 10kb plasmid UUCG1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, P-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gi14732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."  
BASE COUNT 205 a 140 c 108 g 184 t  
ORIGIN  
Query Match 0.6%; Score 21; DB 17; Length 637;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 619 ACTTCAGCTTCATATGAA 639  
|||||  
DB 385 ACTTCAGCTTCATATGAA 365  
|||||  
RESULT 45  
BH098015 755 bp DNA linear GSS 19-JUL-2001  
LOCUS RPI-24-228N6.TJ RPI-24 Mus musculus genomic clone RPI-24-228N6,  
DNA sequence.  
DEFINITION BH098015  
ACCESSION BH098015 GI:14919361  
VERSION BH098015.1  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 755)

**AUTHORS**  
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,  
Tsegaye, G., Geer, K., Kroll, M., Shvartsbeyn, A., Gebregeorgis, E.,  
Russell, D., de Jong, P. and Fraser, C.M.  
**TITLE**  
Mouse BAC End Sequences from Library RPI-24  
**JOURNAL**  
Unpublished (1999)  
**COMMENT**  
Other GSSs: RPI-24-228N6.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPI-24. For BAC  
library availability, please contact Pieter de Jong  
(pdejong@mail.cho.org). Clones may be purchased from BACPAC  
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end  
page: http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html  
Plate: 228 row: N column: 6  
Seq primer: SP6  
Class: BAC ends.  
**FEATURES**  
Location/Qualifiers  
1. 755  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPI-24-228N6"  
/clone\_lib="RPI-24"  
/sex="Male"  
/cell\_type="Spleen/Brain"  
/note="Vector: pTARBAC1; Site\_1: BamHI; Site\_2: BamHI;  
RPI-24 Mouse BAC Library produced by Pieter de Jong. The  
library was cloned in the pTARBAC1 cloning vector at the  
BamHI sites using MboI partially digested male C57BL/6J  
DNA."  
BASE COUNT 175 a 181 c 137 g 262 t  
ORIGIN  
Query Match 0.6%; Score 21; DB 17; Length 755;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 579 AATTTTTCCTTTTATCTA 599  
|||||  
DB 144 AATTTTTCCTTTTATCTA 164  
|||||  
RESULT 46  
BH608326/c 765 bp DNA linear GSS 15-DEC-2001  
LOCUS BOGYJ73TF BOGY Brassica oleracea genomic clone BOGYJ73, DNA  
sequence.  
DEFINITION BH608326  
ACCESSION BH608326 GI:17860772  
VERSION BH608326.1  
KEYWORDS GSS.  
SOURCE Brassica oleracea.  
ORGANISM Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 765)  
Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Other GSSs: BOGYJ73TR  
Contact: Chris Town  
Tigr  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TF



**TITLE**  
Development of a genetically and physically anchored EST resource  
for barley genomics: Morex cold-stressed seedling shoot cDNA  
library

**JOURNAL**  
Unpublished (2001)

**COMMENT**  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
total hq bases = 201  
Seq primer: AATTAACCCCTCAATAAGGG  
High quality sequence start: 110  
High quality sequence stop: 461.

**FEATURES**  
source  
1..954  
/organism="Hordeum vulgare"  
/cultivar="Morex"  
/db\_xref="taxon:4513"  
/clone.lib="HVSMEa0006X07f"  
/clone.lib="Hordeum vulgare seedling shoot EST library  
HVCDA0001 (Cold stress)"  
/tissue\_type="Seedling shoot"  
/lab\_host="TJC121"  
/notes="Vector: lambdaZAP; Site\_1: EcoRI; Site\_2: XhoI;  
Seeds were surface sterilized then germinated under axenic  
conditions in the dark at room temperature on filter paper  
with water, nystatin and cefotaxime in covered  
crystallization dishes. Five-day old seedlings were  
incubated at 50C for 2 days. Shoots were then harvested,  
total RNA was prepared, poly(A) RNA was purified, one  
primary unamplified cDNA library was made, and 600000 pfu  
were in vivo excised to give pBluescript SK(-) cDNA  
phagemids. These steps were performed in the TJ Close  
laboratory at the University of California, Riverside  
(Choi, Close, Fenton). Phagemids were plated and picked at  
the Clemson University Genomics Institute (CUGI) (Begum,  
Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations  
, DNA sequencing and sequence analysis were performed at  
CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main  
). The sequence has been trimmed to remove vector sequence  
and contains a minimum of 100 bases of phred value 20 or  
above. For more details on library preparation and  
sequence analysis see  
http://www.genome.clemson.edu/projects/barley. To order  
this clone see http://www.genome.clemson.edu/orders Also  
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)  
Genetically and physically anchored EST resources for  
barley genomics. Barley Genetics Newsletter 31:29-30.  
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

**BASE COUNT** 166 a 248 c 213 g 325 t 2 others

**ORIGIN**  
Query Match 0.6%; Score 21; DB 12; Length 954;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 AATTTTCTTTCTTTTATCT 598  
|||||  
Db 871 AATTTTCTTTCTTTTATCT 891

**RESULT 50**  
CNS02GIX/c  
LOCUS  
DEFINITION  
Tetraodon nigroviridis genome survey sequence PUC-ori end of clone  
137B20 of library G from Tetraodon nigroviridis, genomic survey  
sequence.  
AL196386  
VERSION AL196386.1 GI:7834536  
KEYWORDS GSS: genome survey sequence.  
SOURCE Tetraodon nigroviridis.

**ORGANISM**  
Tetraodon nigroviridis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodon.  
1 (bases 1 to 1009)  
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.  
Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence  
Unpublished  
2 (bases 1 to 1009)  
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,  
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
Weissenbach,J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish tetraodon nigroviridis  
Unpublished  
3 (bases 1 to 1009)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000)  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetraodon.  
Location/Qualifiers  
1..1009  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone.lib="B20"  
/note="Genoscope sequence ID : COAG137DA10SP1-end :  
PUC-ori"

**BASE COUNT** 359 a 172 c 149 g 326 t 3 others

**ORIGIN**  
Query Match 0.6%; Score 21; DB 17; Length 1009;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 582 TTTTTCCTTTTATCTAGTG 602  
|||||  
Db 238 TTTTTCCTTTTATCTAGTG 218

Search completed: November 6, 2002, 03:25:53  
Job time : 4579 secs





FEATURES		Location/Qualifiers	
source	1..3731	/organism="unknown"	
BASE COUNT	1064 a 669 c 849 g 1149 t		
ORIGIN			
Query Match 100.0%; Score 3731; DB 6; Length 3731; Best Local Similarity 100.0%; Pred. No. 0; Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	GAATATTTCTTACTTACTTACCAGGGATGACTCATCTACCTCCCTCAAGTCTTTGTGT	60
Db	1	GAATATTTCTTACTTACTTACCAGGGATGACTCATCTACCTCCCTCAAGTCTTTGTGT	60
Qy	61	GTGTGATAATAAATTTGGTTGCTTCAGTTTCAGTCACACTGCTGCGGTAGTTTTAT	120
Db	61	GTGTGATAATAAATTTGGTTGCTTCAGTTTCAGTCACACTGCTGCGGTAGTTTTAT	120
Qy	121	TTTGCAATACTTCAGGGGGTATTCAGTTGGTTAGCATTTGAAAGTCGAACTGCACAT	180
Db	121	TTTGCAATACTTCAGGGGGTATTCAGTTGGTTAGCATTTGAAAGTCGAACTGCACAT	180
Qy	181	GGAAATTTGGCTACATGGGAAGACAAATTCAGTTTTCGGATTCCTTATCTCTCTGC	240
Db	181	GGAAATTTGGCTACATGGGAAGACAAATTCAGTTTTCGGATTCCTTATCTCTCTGC	240
Qy	241	GGAAAGTGGTTAAGTCATCTTAGAGAAATATACAGGATATGGAACATGTATGTCATTGGA	300
Db	241	GGAAAGTGGTTAAGTCATCTTAGAGAAATATACAGGATATGGAACATGTATGTCATTGGA	300
Qy	301	GGTTAAACAGTCCAAAGGAGGATCTAGAGCATTTGCCAAAGTTCATTTGCCGACAACT	360
Db	301	GGTTAAACAGTCCAAAGGAGGATCTAGAGCATTTGCCAAAGTTCATTTGCCGACAACT	360
Qy	361	AAAGTGGTACAAAATCATCTTGGCTAATAACAGGCTGTATTTGGCTCTCTTATTT	420
Db	361	AAAGTGGTACAAAATCATCTTGGCTAATAACAGGCTGTATTTGGCTCTCTTATTT	420
Qy	421	GAAGGCTTTGGAAATGAAGAACTGATATGTCCTCAACTCGGGCATATGTCGATGGA	480
Db	421	GAAGGCTTTGGAAATGAAGAACTGATATGTCCTCAACTCGGGCATATGTCGATGGA	480
Qy	481	TGSCATAACTTTGAATTCGGATGTCAGATATCAGATGACAAGTTTGCAGTGTGGGAAG	540
Db	481	TGSCATAACTTTGAATTCGGATGTCAGATATCAGATGACAAGTTTGCAGTGTGGGAAG	540
Qy	541	TACAGAACTTTCAATTTGGCAATTTGGCAATTTTGGCAATTTTCTTTTATCTAG	600
Db	541	TACAGAACTTTCAATTTGGCAATTTGGCAATTTTGGCAATTTTCTTTTATCTAG	600
Qy	601	TGGTTACGCTGACTATAAATTCAGCTTCATATGAAATATATGCGAGTGTGCTCCA	660
Db	601	TGGTTACGCTGACTATAAATTCAGCTTCATATGAAATATATGCGAGTGTGCTCCA	660
Qy	661	TCGTCCATATGGTCAAAATGCTCAGTTTCCCTCATACAGTTATTTGGTCTCTCGGAT	720
Db	661	TCGTCCATATGGTCAAAATGCTCAGTTTCCCTCATACAGTTATTTGGTCTCTCGGAT	720
Qy	721	CTATAAGAGACTTGAAGAACTCCCTGTTATAGCTTCTTTAAGGAACTCCCTGATCAGTG	780
Db	721	CTATAAGAGACTTGAAGAACTCCCTGTTATAGCTTCTTTAAGGAACTCCCTGATCAGTG	780
Qy	781	GGTGAGGACACAGATTTCCCTCATCTTGGATAGGCTATCTTCTAGCTTATGTTTGA	840
Db	781	GGTGAGGACACAGATTTCCCTCATCTTGGATAGGCTATCTTCTAGCTTATGTTTGA	840
Qy	841	GTTCGGTAGGGGTTCGCTTCTTCCAAATTCGAGGAAAGTTTTTCACATATCGCAAGC	900
Db	841	GTTCGGTAGGGGTTCGCTTCTTCCAAATTCGAGGAAAGTTTTTCACATATCGCAAGC	900
Qy	901	TGAACAAATATTAATTTACAGACTGGTTTCACTTTTTCGCTCTCTCAAAAATTCGGCTCT	960
Db	901	TGAACAAATATTAATTTACAGACTGGTTTCACTTTTTCGCTCTCTCAAAAATTCGGCTCT	960

Qy	961	GGTTCCCAATGTCCAGCCTCCGGAAGAAATTTCAATTCCTACAGATTTTGTCAAAAT	1020
Db	961	GGTTCCCAATGTCCAGCCTCCGGAAGAAATTTCAATTCCTACAGATTTTGTCAAAAT	1020
Qy	1021	TAGTTCTTTGGTACAGCATGGATGATACCTGGGCCAGCATTAATGTCTACTTTTCGG	1080
Db	1021	TAGTTCTTTGGTACAGCATGGATGATACCTGGGCCAGCATTAATGTCTACTTTTCGG	1080
Qy	1081	ATTAGTTGATCCCTCGAAGAGAGAAATGTGGCATGCAATGAGCATGCCCTTAGAGAACTGTA	1140
Db	1081	ATTAGTTGATCCCTCGAAGAGAGAAATGTGGCATGCAATGAGCATGCCCTTAGAGAACTGTA	1140
Qy	1141	CTATATAAAGGAGTGTGTATGATCCCGTGAGTGGCTCACTGAGCAGTATGATGGGTA	1200
Db	1141	CTATATAAAGGAGTGTGTATGATCCCGTGAGTGGCTCACTGAGCAGTATGATGGGTA	1200
Qy	1201	TCTCAAGGGTACACAACCTCCAAAATCTCGCTCCATCACTTTAGATGATGGGTGGTGA	1260
Db	1201	TCTCAAGGGTACACAACCTCCAAAATCTCGCTCCATCACTTTAGATGATGGGTGGTGA	1260
Qy	1261	TGTAAAGGGTCCCTAGTAAACACCATGCAAGTATTATTTTGTGGTCCAGAGTTAATGT	1320
Db	1261	TGTAAAGGGTCCCTAGTAAACACCATGCAAGTATTATTTTGTGGTCCAGAGTTAATGT	1320
Qy	1321	TTCCAATCGGGTTCCTCCGCAATTTATCTGAAGACATAGATAACTTCTCTGTTCTTT	1380
Db	1321	TTCCAATCGGGTTCCTCCGCAATTTATCTGAAGACATAGATAACTTCTCTGTTCTTT	1380
Qy	1381	TGTTGATGAGGAGTGGGAGAACTGTATTTCTACAGACTTATACCAAAGCAAGTACTGG	1440
Db	1381	TGTTGATGAGGAGTGGGAGAACTGTATTTCTACAGACTTATACCAAAGCAAGTACTGG	1440
Qy	1441	AAAGTGGTGCAGACAAACATCTATGAGAGGATCTTATCAACTCTCGGAAAGGCTTGT	1500
Db	1441	AAAGTGGTGCAGACAAACATCTATGAGAGGATCTTATCAACTCTCGGAAAGGCTTGT	1500
Qy	1501	AATTTGATATAAAATTTGAATTTCTTGCATTTTCATCGAGCCAGTTGCGGGAATAATC	1560
Db	1501	AATTTGATATAAAATTTGAATTTCTTGCATTTTCATCGAGCCAGTTGCGGGAATAATC	1560
Qy	1561	AGTGTGGATTTTGCATCAAGACCTGGCCTTACTCAATGATATAAGAGCTTGGATGG	1620
Db	1561	AGTGTGGATTTTGCATCAAGACCTGGCCTTACTCAATGATATAAGAGCTTGGATGG	1620
Qy	1621	TGATTTTTCGCAGATCAAGAAATGTGCAAAATATGCTGCCAGACTTGGTCAATCTTTGG	1680
Db	1621	TGATTTTTCGCAGATCAAGAAATGTGCAAAATATGCTGCCAGACTTGGTCAATCTTTGG	1680
Qy	1681	TTCCCTCCAGAGAGCTTTGAGTGTCTTAGGATGAGATGAAGTATTCGGATGTAAA	1740
Db	1681	TTCCCTCCAGAGAGCTTTGAGTGTCTTAGGATGAGATGAAGTATTCGGATGTAAA	1740
Qy	1741	GGTTATGGAACAGCTATGCTTTTCTGATGGAATTTGTAATAATCTGTTGACTTTGC	1800
Db	1741	GGTTATGGAACAGCTATGCTTTTCTGATGGAATTTGTAATAATCTGTTGACTTTGC	1800
Qy	1801	TCATAGAGTTGCTCAAAATGTGGCTTCAATATACCCCATCTGCTTTCCAGATTCGTA	1860
Db	1801	TCATAGAGTTGCTCAAAATGTGGCTTCAATATACCCCATCTGCTTTCCAGATTCGTA	1860
Qy	1861	TGGTGGATATAAAGGTGTGTGGGTGTGATCCGGATTCATCAATGAAGTGTCTTTGAG	1920
Db	1861	TGGTGGATATAAAGGTGTGTGGGTGTGATCCGGATTCATCAATGAAGTGTCTTTGAG	1920
Qy	1921	AAAGAGCATGTCGAATATGAATCAGACACATATAAGTTAGATGCTTGGATGAGCAA	1980
Db	1921	AAAGAGCATGTCGAATATGAATCAGACACATATAAGTTAGATGCTTGGATGAGCAA	1980
Qy	1981	ATATCAGCCTTCTTATCTTAAATCGTCAACTGATTAACGCTTCTTGTCTACACTTGGAGTAA	2040
Db	1981	ATATCAGCCTTCTTATCTTAAATCGTCAACTGATTAACGCTTCTTGTCTACACTTGGAGTAA	2040



/organism="Lycopersicon esculentum"			
/strain="Rutgers"			
/db_xref="taxon:4081"			
gene	194...3538		
CDS	/gene="RdRp"		
	194...3538		
	/gene="RdRp"		
	/EC_number="2.7.7.48"		
	/codon_start=1		
	/product="RNA-directed RNA polymerase"		
	/protein_id="CAA71421.1"		
	/db_xref="GI:4038592"		
/translation="MGKTTQVGFPGYLLSAEVVSKLEKYGTGTVCALEVKQSKGGS RAFAQVQADNISADRIILIANRLNRFSGSYLKAWEKTDIVQLRAYVQDMGDTLNF CGOISDDKFAVLGSTEVSIOFGILKLFKFFFLSSGSADYKQLQSYENIQWVYLHRYG ONAFLLIQFAGPRVYKRLNCSYGFETPDQWRTTDRPPSYGLSSSLCLQFPR RGVLPNFESEFFPHYAERENNIITLQGETFFVSOKSALVNPVQPPGIIIPYKILFKI SSLVGHGCIIPGPAHYVEFRLVDPBRNVACIEHLEKLYIKKCCYDVRVRLTEGYD GIYKGRQPKSPSITLDDGLVYVRVLYVPCFYFCGPEVNSVRNIRLSEIDINFL RVSEVDEWEKLYSTDLLPKASTGSGVRINIERILSTLRKGFVIGDKKPELAFSSS QLRDSNVMFASRPGLTANDIRAMGDFSOIKNVAKYARLQSGFSSRETLSVLRHE IEVIPDKVHGTSYVFSDGIGKISGFHVRVSKGLQYTPSAFOIRYGGYGVVGD PDSMKLSLKSNSKYESNLIKLDVLGWSKYQPCYNLRQILITLLSLGYKDEVLQKQ KEAVDOLDALHDSLAQEALEMLSPGENTILKMLNGCYKPDAPPEFSLMLQFRA SKLLDLRTSRRIPIPNGRPMWGLDERSLEYGOVFQVQFTGAGHGESDILHPRNRS STNSNTILKGNVVAKNPCILHPGDIRVLKXNVNRAHLHMVDCVFPQKRRHPNECS GSDLDGIYFCWQDMKIPROVQPMEXPPAPSIOLDHDVYIIEVEEYETNFIYNDLS GIIANHVVFADEPDMAMSDPCKLAEFLSAVDFFPGTGVFAEIPSQLRPKYEYDFM ADPKTSYISERVIGKLFKRYDKDKPAQSSIAFTFRDVARRYSDADMEYDGFEDIDE AFDYKTEYDNKGNLDYIGIKTEAEILSGGIMKASKTFDRRDKDAEISAVYARLKE ARAWFKRRNDIDMLPKASAWYHVYHTPYTGWGCYNQGLKRAHFISFPWCYVDQLIQIK KDKARNRPNLNLSSLAQLSHRLVK"			
BASE COUNT	1064 a 668 c 850 g 1149 t		
ORIGIN			
Query Match 100.0%; Score 3729.4; DB 8; Length 3731;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 3730; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
QY	1	GAATATCTTTACTTACTTACCAGGATGTGACTCATCTACCTCCCTCAAGTCTTTGTG 60	
DB	1	GAATATCTTTACTTACTTACCAGGATGTGACTCATCTACCTCCCTCAAGTCTTTGTG 60	
QY	61	GTGTGATAATAATTTGGTGTGCTTCAGTTTCAGTCACTACTGCTGGGTAGTCTTTAT 120	
DB	61	GTGTGATAATAAATTTGGTGTGCTTCAGTTTCAGTCACTACTGCTGGGTAGTCTTTAT 120	
QY	121	TTTGCATAAATTCAGGGGTATCCAGTTGGTGTAGCATTTGAAAGTCGAAGTGCACIT 180	
DB	121	TTTGCATAAATTCAGGGGTATCCAGTTGGTGTAGCATTTGAAAGTCGAAGTGCACIT 180	
QY	181	GGAATTTGGCTACATGGGAAAGACAATTCAGGTTTTCGGATTCCTTATCTCTCTGC 240	
DB	181	GGAATTTGGCTACATGGGAAAGACAATTCAGGTTTTCGGATTCCTTATCTCTCTGC 240	
QY	241	GGAAGTGGTTAAGTCATCTTTAGAGAAATATACAGGATATGGAAGTGTATGTGCATTTGA 300	
DB	241	GGAAGTGGTTAAGTCATCTTTAGAGAAATATACAGGATATGGAAGTGTATGTGCATTTGA 300	
QY	301	GGTTAAACAGTCCAAAGAGGATCTAGAGCAATTTGCCAAAGTTCAATTTGCCGACACAAT 360	
DB	301	GGTTAAACAGTCCAAAGAGGATCTAGAGCAATTTGCCAAAGTTCAATTTGCCGACACAAT 360	
QY	361	AAGTGTGACAAAATCATCTATTTGGCTAATAACAGGCTGTATTTTGGCTCTCTTATTT 420	
DB	361	AAGTGTGACAAAATCATCTATTTGGCTAATAACAGGCTGTATTTTGGCTCTCTTATTT 420	
QY	421	GAAGGCTTGGGAATGAAACTGATATTTGCCAACTCGGGGCATATGTGCATCAGATGGA 480	
DB	421	GAAGGCTTGGGAATGAAACTGATATTTGCCAACTCGGGGCATATGTGCATCAGATGGA 480	
QY	481	TGGCATAACTTTGAATTTCCGATGTCAGATATCAGATGACAAGTTTCAGTGTGGGAAG 540	
DB	481	TGGCATAACTTTGAATTTCCGATGTCAGATATCAGATGACAAGTTTCAGTGTGGGAAG 540	
QY	541	TACAGAAAGTTTCAATTCATTTGGCATTTGGATGAAGAAATTTTTTTTCTTTTATCTAG 600	
DB	541	TACAGAAAGTTTCAATTCATTTGGCATTTGGATGAAGAAATTTTTTTTCTTTTATCTAG 600	
QY	601	TGGTTACAGCTGACTATAAAACTTCAGCTTTCATATGAAAATATATGCGCAGGTTGTGCTCCA 660	
DB	601	TGGTTACAGCTGACTATAAAACTTCAGCTTTCATATGAAAATATATGCGCAGGTTGTGCTCCA 660	
QY	661	TGTCCTCATATGTTGTCAAAATGCTCAGTTTCTCTCATACAGTTTATTTGGTGTCTCTCGGAT 720	
DB	661	TGTCCTCATATGTTGTCAAAATGCTCAGTTTCTCTCATACAGTTTATTTGGTGTCTCTCGGAT 720	
QY	721	CTATAAGAGACTTGAAAACCTCCTGTTTATAGTCTTTTAAAGGAAACCTCCTGATGATCAGTG 780	
DB	721	CTATAAGAGACTTGAAAACCTCCTGTTTATAGTCTTTTAAAGGAAACCTCCTGATGATCAGTG 780	
QY	781	GGTGAGGACAAACAGATTTCCCTCCATCTTTGGATAGGGCTATCTTCTAGCTTATGTTTGA 840	
DB	781	GGTGAGGACAAACAGATTTCCCTCCATCTTTGGATAGGGCTATCTTCTAGCTTATGTTTGA 840	
QY	841	GTTCCTAGGGGTGTTGCTCTTCCAAATTCGAGGAAAGTTTTTTCCACTATGCGAAGCG 900	
DB	841	GTTCCTAGGGGTGTTGCTCTTCCAAATTCGAGGAAAGTTTTTTTCCACTATGCGAAGCG 900	
QY	901	TGAAAACAATATTAATTTACAGACTGGTTTTCACCTTTTTTTCGCTCTCTCAAAAATCGGCTCT 960	
DB	901	TGAAAACAATATTAATTTACAGACTGGTTTTCACCTTTTTTTCGCTCTCTCAAAAATCGGCTCT 960	
QY	961	GGTTCCCAATGTCACGCTCCGGAAGGAATTTCAATTCCTTCAAGATTTTGTTCAAAAT 1020	
DB	961	GGTTCCCAATGTCACGCTCCGGAAGGAATTTCAATTCCTTCAAGATTTTGTTCAAAAT 1020	
QY	1021	TAGTTCTTTGGTACAGCATGGATGCATACCTGGCCAGCAATTAATGTCTACTTTTTCCG 1080	
DB	1021	TAGTTCTTTGGTACAGCATGGATGCATACCTGGCCAGCAATTAATGTCTACTTTTTCCG 1080	
QY	1081	ATTAGTTGATCCTCGAAGGAAATGTGGCATGTCATGAGCATGCCCTTAGAGAAACTGTA 1140	
DB	1081	ATTAGTTGATCCTCGAAGGAAATGTGGCATGTCATGAGCATGCCCTTAGAGAAACTGTA 1140	
QY	1141	CTATATAAGAGAGTGTCTTTATGATCCCGTGGGTGGCTCACTGAGCAGTATGATGGGTA 1200	
DB	1141	CTATATAAGAGAGTGTCTTTATGATCCGGTGGGTGGCTCACTGAGCAGTATGATGGGTA 1200	
QY	1201	TCTCAAGGTTAGACAACTCCAAAATCTCCGTCCTCATCACTTTAGATGATGGGTTGGTGA 1260	
DB	1201	TCTCAAGGTTAGACAACTCCAAAATCTCCGTCCTCATCACTTTAGATGATGGGTTGGTGA 1260	
QY	1261	TGTAAGAAGGGTCTCTAGTAACACCATGCAAAAGTTTATTTTGTGGTCCAGAGGTTAATGT 1320	
DB	1261	TGTAAGAAGGGTCTCTAGTAACACCATGCAAAAGTTTATTTTGTGGTCCAGAGGTTAATGT 1320	
QY	1321	TTCCAATCGGGTCTCCGCAATTAATCTTGAAGACATAGATAACTTTCTTCGTGTTTCTTT 1380	
DB	1321	TTCCAATCGGGTCTCCGCAATTAATCTTGAAGACATAGATAACTTTCTTCGTGTTTCTTT 1380	
QY	1381	TGTTGATGAGGAGTGGGAGAACTGTATTTCTACAGCTTATTTACCAAAAGCAAGTACTGG 1440	
DB	1381	TGTTGATGAGGAGTGGGAGAACTGTATTTCTACAGACTTATTTACCAAAAGCAAGTACTGG 1440	
QY	1441	AAGTGGTGTGACGACAAACATCTATGAGAGGATCTTATCAACTCTCGGGAAGGCTTTGT 1500	
DB	1441	AAGTGGTGTGACGACAAACATCTATGAGAGGATCTTATCAACTCTCGGGAAGGCTTTGT 1500	
QY	1501	AATTGGTGATAAAAAATTTGAATTTCTTGCATTTTTCATCGAGCCAGTTTGGGGATAATTC 1560	
DB	1501	AATTGGTGATAAAAAATTTGAATTTCTTGCATTTTTCATCGAGCCAGTTTGGGGATAATTC 1560	
QY	1561	AGTGTGGATTTTGCATCAAGACCTGGCCTTACTGCAATGATATAAGAGCTTGGATGGG 1620	
DB	1561	AGTGTGGATTTTGCATCAAGACCTGGCCTTACTGCAATGATATAAGAGCTTGGATGGG 1620	

QY 1621 TGATTTTTCGAGATCAAGATGTCGCAAAATATGCTGCCAGACTTGGTCAATCTTTGG 1680  
DB 1621 TGATTTTTCGAGATCAAGATGTCGCAAAATATGCTGCCAGACTTGGTCAATCTTTGG 1680  
QY 1681 TTCCTCCAGAGAGACTTTGAGTGTCTTTAGGCATGAGATTGAAGTTATTCGCCGATGAAA 1740  
DB 1681 TTCCTCCAGAGAGACTTTGAGTGTCTTTAGGCATGAGATTGAAGTTATTCGCCGATGAAA 1740  
QY 1741 GGTTCATGGAAACCAAGCTATGCTTTTTCGATGGAATGGTAAATATCTGGTGCATTTGC 1800  
DB 1741 GGTTCATGGAAACCAAGCTATGCTTTTTCGATGGAATGGTAAATATCTGGTGCATTTGC 1800  
QY 1801 TCATAGAGTTCGCTCAAAATGTCGCTCAATATACCCATCTCTCTTCCAGATTCGTTA 1860  
DB 1801 TCATAGAGTTCGCTCAAAATGTCGCTCAATATACCCATCTCTCTTCCAGATTCGTTA 1860  
QY 1861 TGGTGGATATAAGGTGTGTGGTGTGTGATCCGGATTCATCAATGAAGTTGTCTTTGAG 1920  
DB 1861 TGGTGGATATAAGGTGTGTGGTGTGTGATCCGGATTCATCAATGAAGTTGTCTTTGAG 1920  
QY 1921 AAAGGAGTTCGCAAAATGATGAATCAGACACATAAAGTTAGATGCTCTTGGATGGACAA 1980  
DB 1921 AAAGGAGTTCGCAAAATGATGAATCAGACACATAAAGTTAGATGCTCTTGGATGGACAA 1980  
QY 1981 ATATCAGCCTTGTATCTTAACTCGTCAACTGATTACGCTCTTGTCTACACTTGGAGTGA 2040  
DB 1981 ATATCAGCCTTGTATCTTAACTCGTCAACTGATTACGCTCTTGTCTACACTTGGAGTGA 2040  
QY 2041 AGATGAAGTTCGCAACAGAAAGAAAGAGCTGTAGATCAGCTTGATGCTATCTTGCAT 2100  
DB 2041 AGATGAAGTTCGCAACAGAAAGAAAGAGCTGTAGATCAGCTTGATGCTATCTTGCAT 2100  
QY 2101 TGATCTTTTGAAGCACAGGAGGCTTTGGAAATGATGCTCTCTGGAGAGAACACTAATAT 2160  
DB 2101 TGATCTTTTGAAGCACAGGAGGCTTTGGAAATGATGCTCTCTGGAGAGAACACTAATAT 2160  
QY 2161 TCTCAAGCAATGCTAACTGTGGTTATAAGCCGTGATGCTGAGCCCTTTCTTCAATGAT 2220  
DB 2161 TCTCAAGCAATGCTAACTGTGGTTATAAGCCGTGATGCTGAGCCCTTTCTTCAATGAT 2220  
QY 2221 GTTGCAAACCTTCGCGCATCCAAAGTTGCTCGATTTCGGAGCTAGATCAAGATATTTAT 2280  
DB 2221 GTTGCAAACCTTCGCGCATCCAAAGTTGCTCGATTTCGGAGCTAGATCAAGATATTTAT 2280  
QY 2281 TCCAAATGGAAGAACATGATGGATGTTTGGATGAATCCAGAACCTTGGAAATATGGTCA 2340  
DB 2281 TCCAAATGGAAGAACATGATGGATGTTTGGATGAATCCAGAACCTTGGAAATATGGTCA 2340  
QY 2341 GGTCTTTGCTCAGTTACTGGTGTGCACATGGAGATTTCTGACGATTTACATCCATT 2400  
DB 2341 GGTCTTTGCTCAGTTACTGGTGTGCACATGGAGATTTCTGACGATTTACATCCATT 2400  
QY 2401 TAATAACAGCAGATCCACCAACAGTAAATTCATTCTGAAGGAAATGTGGTTGTGCAAA 2460  
DB 2401 TAATAACAGCAGATCCACCAACAGTAAATTCATTCTGAAGGAAATGTGGTTGTGCAAA 2460  
QY 2461 AAATCCATGCTGCATCCTGGTGATATTCGTGTTTAAAGCGTGTAATGTTTCGAGCGCT 2520  
DB 2461 AAATCCATGCTGCATCCTGGTGATATTCGTGTTTAAAGCGTGTAATGTTTCGAGCGCT 2520  
QY 2521 GCACCATGCTGATGTTGTTGTTTCCCTCAGAAAGGAAAGAACCTCATCCGAATGA 2580  
DB 2521 GCACCATGCTGATGTTGTTGTTTCCCTCAGAAAGGAAAGAACCTCATCCGAATGA 2580  
QY 2581 ATGCTCGGAGTGAATTTGGATGGGATATCTACTTTGTTGCTGGGATCAAGACATGAT 2640  
DB 2581 ATGCTCGGAGTGAATTTGGATGGGATATCTACTTTGTTGCTGGGATCAAGACATGAT 2640  
QY 2641 CCGCCCAAGCAAGTCCAGCCGATGGATATATCTCAGACCCAGCATACAGTTGGACCA 2700  
DB 2641 CCGCCCAAGCAAGTCCAGCCGATGGATATATCTCAGACCCAGCATACAGTTGGACCA 2700

QY 2701 TGATGTCACAATTCAGGAAGTTGAAGAGTACTTCCACCAACTATATTGTGAATGACAGTTT 2760  
DB 2701 TGATGTCACAATTCAGGAAGTTGAAGAGTACTTCCACCAACTATATTGTGAATGACAGTTT 2760  
QY 2761 GCGAATCATAGCAAAATGCCCATGCTGATTTGCGAGACAGAACCTGATATGCCATGAG 2820  
DB 2761 GCGAATCATAGCAAAATGCCCATGCTGATTTGCGAGACAGAACCTGATATGCCATGAG 2820  
QY 2821 TGATCCATGCAAAAAACTTGTCTGAGCTCTTTTCAATTCAGTGGAGCTTTTCCAAAGACTGG 2880  
DB 2821 TGATCCATGCAAAAAACTTGTCTGAGCTCTTTTCAATTCAGTGGAGCTTTTCCAAAGACTGG 2880  
QY 2881 TGTTCCTCGTGAATACCATCTCAGTTCGCCCCCTAAAGAAATACCCAGACTTCATGATATA 2940  
DB 2881 TGTTCCTCGTGAATACCATCTCAGTTCGCCCCCTAAAGAAATACCCAGACTTCATGATATA 2940  
QY 2941 GCCGGCAAGACCAAGCTATATCTCAGAAAGAGTTATTGGAAGCTTTTTCAGGAAAGTGAA 3000  
DB 2941 GCCGGCAAGACCAAGCTATATCTCAGAAAGAGTTATTGGAAGCTTTTTCAGGAAAGTGAA 3000  
QY 3001 GGCAAAAGCACCTCAGGCTAGCTCTATCGGGACCTTCAAGAGATGTTGCAAGGAGATC 3060  
DB 3001 GGCAAAAGCACCTCAGGCTAGCTCTATCGGGACCTTCAAGAGATGTTGCAAGGAGATC 3060  
QY 3061 ATATGATGCTGATATGGAAGTTGATGATTTGAAGATTACATTGACGAAAGCTTTTGACTA 3120  
DB 3061 ATATGATGCTGATATGGAAGTTGATGATTTGAAGATTACATTGACGAAAGCTTTTGACTA 3120  
QY 3121 CAAACTGATATGCAACAAGCTGGTAAATTAATGGACTACTATGGCATATAAAACAGA 3180  
DB 3121 CAAACTGATATGCAACAAGCTGGTAAATTAATGGACTACTATGGCATATAAAACAGA 3180  
QY 3181 GCGTGAATACTAGTGGTGGCATTAAGGCATCAAAACTTTTGACCGCAGAAAGA 3240  
DB 3181 GCGTGAATACTAGTGGTGGCATTAAGGCATCAAAACTTTTGACCGCAGAAAGA 3240  
QY 3241 TGTGTAGGCCATTAGTGTGCTGTGAGGCCCTTGGAGAGGAGGCAAGACCTTGTTCAA 3300  
DB 3241 TGTGTAGGCCATTAGTGTGCTGTGAGGCCCTTGGAGAGGAGGCAAGACCTTGTTCAA 3300  
QY 3301 GAGGCTTAATGATATAGATGACATGTTACCAAGGCTTCGGCTTGTACCAAGCTTACATA 3360  
DB 3301 GAGGCTTAATGATATAGATGACATGTTACCAAGGCTTCGGCTTGTACCAAGCTTACATA 3360  
QY 3361 TCATCTACATATTGGGGTGTCTACAAATCAGGGGTTGAAAAGAGCTCATTTCAATAGCTT 3420  
DB 3361 TCATCTACATATTGGGGTGTCTACAAATCAGGGGTTGAAAAGAGCTCATTTCAATAGCTT 3420  
QY 3421 TCCCTGGTGTGTTATGACCAGCTAATCCAGATTAAAGAGGACAAAGACGTAACAGGCC 3480  
DB 3421 TCCCTGGTGTGTTATGACCAGCTAATCCAGATTAAAGAGGACAAAGACGTAACAGGCC 3480  
QY 3481 AGTTCTCAACTTGTCTATCTCTCAGGGCTCAACTGAGTCAAGATTAGTGTGAAATGAGA 3540  
DB 3481 AGTTCTCAACTTGTCTATCTCTCAGGGCTCAACTGAGTCAAGATTAGTGTGAAATGAGA 3540  
QY 3541 TTCAGTCCAGGGTAAAGCTGATATATAATGTAATAGGGTGTGATCATAAAGAAAC 3600  
DB 3541 TTCAGTCCAGGGTAAAGCTGATATATAATGTAATAGGGTGTGATCATAAAGAAAC 3600  
QY 3601 TGTGTGACATTTGACTACCTTTTGTCTTTAAACTGCAAGCTGCAACATATATGC 3660  
DB 3601 TGTGTGACATTTGACTACCTTTTGTCTTTAAACTGCAAGCTGCAACATATATGC 3660  
QY 3661 AGTACTCTAAGAAACAGATGATACAGTAACTAACTAACTAACTAACTAACTAACTAACT 3720  
DB 3661 AGTACTCTAAGAAACAGATGATACAGTAACTAACTAACTAACTAACTAACTAACTAACT 3720  
QY 3721 TTTCTCTTAAA 3731  
DB 3721 TTTCTCTTAAA 3731

```

RESULT 3
NTA011576      3505 bp      mRNA      linear      PLN 07-JAN-1999
LOCUS          Nicotiana tabacum mRNA for RNA-directed RNA polymerase.
DEFINITION
ACCESSION     AJ011576
VERSION       AJ011576.1  GI:4138281
KEYWORDS      drp gene; RNA-directed RNA polymerase.
SOURCE        common tobacco.
ORGANISM      Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; eustersids I; Solanales; Solanaceae; Nicotiana.
Wassenegger M.
Direct Submission
Submitted (16-OCT-1998) Wassenegger M., Max-Planck-Institut of
Biochemistry Viroidresearch, Am Klopferspitz 18A, 82152
Martinsried, FRG
2 (bases 1 to 3505)
Schiebel, W., Pelissier, T., Riedel, L., Thalmair, S., Schiebel, R.,
Kempe, D., Lottspeich, F., Saenger, H. L. and Wassenegger M.
Isolation of an RNA-directed RNA polymerase-specific cDNA clone
from Tomato
Unpublished
FEATURES             Location/Qualifiers
     source           1..3505
                     /organism="Nicotiana tabacum"
                     /cultivar="Petit Havana SRI"
                     /db_xref="taxon:4097"
     gene             20..3370
                     /tissue_type="leaves"
     CDS              20..3370
                     /gene="RdRP"
                     /codon_start=1
                     /evidence=experimental
                     /product="RNA-directed RNA polymerase"
                     /protein_id="CAA09697.1"
                     /db_xref="GI:4138282"
                     /translation="MGKTIQVGFPPYLLSAEAVKFLNHTNGIVYALEVQSKGGR
                     RAFPQVANKSAEFLIDLSKGLYIGSYLKAWEKMDIVOPRTYVEHMDGVTLNF
                     GCQIDKKFAVLGSDVSIKFGILKIKYFFLSHASADYKLQLSYENTWQVYLHRPYG
                     QNAQFLLLQFAGPRIYKRLSDSCYFFKETPDQDQWRTDFTPSWIGLSSLCLEFR
                     NGVLPNFSEFSEFYKESMNQFILOTGFTFSQKALVPIVHLPEGLPKYILKFI
                     SLIOHGCPGLALNENFQVDPRRNACIEHALEKLYLKECCYDPVRLTEQVD
                     EXLKGRLPKSPITLDDGLVYRVVVTVPCKYFCGPPVNVSNVRLNYSEDIINFL
                     RVSFVDEEKEKHSIDLLPRSTNGTRDIYERILSLNRGFIIGDKRFEAFSSS
                     QLRDNSVWFAKPGLTANDIRTMGDFQIRNVAKYARLGQSFSGSRETLISVGHE
                     VEPIDVACSLHGTNYIFSDGIGKISADFAHRTIKCGLQYTPSPFQIRYGGYKQVVA
                     VDPYSMSLSRKSMLKYESNNIKLDVLGSKYQPCYNLRQLVTLISLGVDDVLEQ
                     KONEAYDQLDAIHLSLKAQEALELMSPGENTNILEMLNCGYMPDAPPEFLSMIQTQ
                     RASKLLDLRTRIFIPNGRTWMGCLDESRTLEYGVQVQFSGAGRPFESHPND
                     SCSANDCFILKGNVAVKNCPLHPGDIRVLRADYVPAALHHMYDVCVYFQKGRHPNE
                     CSGSDLDGDIYFCVWDPLIPRQVQSDYTPAPTQLDHDVITIEVEEYFNYIIND
                     SLGIIANHFVADRPDMASDPCKQLAQLFSIAVDPKTGPAPFISQLRPKEYPD
                     FMEKDPKHYPEKVIKGLFKQVKNTPQASSIATFRDVARSKYSDSDMIVDGFEDYI
                     DEAFYKSYDNKGLMDYGIKTEAEILSGIMKASKTFRDRKDAIGAIVAVCLR
                     KEARAFKRSRSDIDMLAKASAWHYVHTYTWGLYNEGLKRDHFIISFPWCYVDLIQ
                     IKKAKARKPEVPVHLSLSGSLSRKLVW"
BASE COUNT      1020 a      622 c      785 g      1078 t
ORIGIN
Query Match      76.5%; Score 2852.8; DB 8; Length 3505;
Best Local Similarity 89.0%; Pred. No. 0;
Matches 3137; Conservative 0; Mismatches 362; Indels 27; Gaps 4;

QY 175 GCACCTTGGAAATTTGGTACATGGAAGACAATTCAGGTTTCGGATTCCCTATCTTCT 234
DB 1 GCACCTTGGAAATTTGGTACATGGAAGACAATTCAGGTTTCGGATTCCCTATCTTCT 60

QY 235 CTCTCGCGGAAGTGGTAAAGTCATTCTAGAGAAATATACAGGATATGGAACTGTATGTGC 294
DB 61 GTCTGCGAGACAGTCAGAGAAATCTTAGAGATCATACAGGAATGGAACTGTCTATGC 120

```



```
Db 3361 GAT-----ATGGTAGCATTAAAGGAGCTATATA-ATGTAAAAAGGGTGTGA 3406
QY 3589 TCATAAGAAACTGTATGCAATGTTGACTACCTTTTCTCTTTAAACTGCATGAAGCTG 3648
Db 3407 TCATAAGAACTGTATGCTGCTGTGTAAACACCTTTTCTCTTTAAACT-----GGTG 3459
QY 3649 CAACATATACGAGTACCTAAGAAACAGATGACAGCTAAGTACT 3694
Db 3460 ATATATTTATGACACCTCTTTAAACAGATGTTATAGCTAAGTAGT 3505

RESULT 4
PHY011979
LOCUS Petunia hybrida Rdrp gene, partial. 4579 bp DNA linear PLN 07-JAN-1999
DEFINITION AJ011979
ACCESSION AJ011979
VERSION AJ011979.1 GI:4138342
KEYWORDS RDRP gene; RNA-directed RNA polymerase.
SOURCE Petunia x hybrida
ORGANISM Petunia x hybrida
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
1 (bases 1 to 4579)
Wassenegger,M.
Direct Submission
Submitted (16-OCT-1998) Wassenegger M., Max-Planck-Institut of
Biochemistry Viroidresearch, Am Klopferspitz 18A, 82152
Martinsried, FRG
2 (bases 1 to 4579)
Schleib,W., Pelissier,T., Riedel,L., Thalmair,S., Schiebel,R.,
Kempe,D., Lottspeich,F., Saenger,H.B. and Wassenegger,M.
Isolation of an RNA-directed RNA polymerase-specific cDNA clone
from Tomato
Unpublished
Location/Qualifiers
1..4579
/organism="Petunia x hybrida"
/db_xref="taxon:4102"
/tissue_type="leaves"
1..4579
/gene="Rdrp"
Join(<1..1451,2420..2570,3856..>4579)
/gene="Rdrp"
/codon_start=1
/evidence-experimental
/product="RNA-directed RNA polymerase"
/protein_id="CAA09896.1"
/db_xref="GI:4138343"
/translators="CIEHALEKLYNKECCYDPRWLTKOYDEYLKRRLPKSPSTIL
DGLMYVRRVLVTPCKVYFLGPEVNVNRLVNYSEDIDNLFVRVDFVDFEWFISMD
LLPRANRGTDTDIYRIILSTLQNLVIGAKKEFLAFSSQLRDSNVMPASRGLT
AYDIRTWGDFQDIRNVAKYARLGOSFSSRETLVGRHEIKYIPDVEVHGTNVYFS
DGKISADFAHRVAKKCGLOHTPSAFQIRYGGYKGVAVDPYSPMKLSRLSMLKYE
SNTKLDLVGWSKYQPCYLNRLQTLMSLTLGVKDEYFKRQNEAVQLDALHDSLKA
QEAALMSFGENTNILEMLNCGYLPDAEPFLSMLOTFRASKLLDLRTRIFIPKG
RSMWGLDRESRTLEYGVVFOFTGAGROFYEEHSHFNDSGANVFIKGNVYAKN
PCHPQDRIYLVKAVDPALHMDVGVVFPQKGRPHNPSGSDLDLDIYFVCDPDL
IPROVQPDYTPAPSIQLDHDTVEEVEEFTNMYNDGLGVIANAHVVPADREPML
AMSDPCLAQLESIAVDFPKGVPAEISQLRPKEYPDFNEKPTTTHSEKVIKGL
FRVKDKTKQASSTITTRDVAHRSYDLEVDGFEDYIDEAFDYKSEYDKLGLMD
YGIKTEAETLSGIMKASKFTDRRDAEIGAIVAVESRLKEARTWFKRSDIDDLIAK
ASAWYHVAYHTYWGVCYNEGLKRDHFISFPWCYVD"
<1..1451
/gene="Rdrp"
/number=1
1452..2419
/gene="Rdrp"
/number=1
2420..2570
/gene="Rdrp"
/number=2
2571..3855

exon
intron
exon
intron
```









Db	55153	TGTAATGGTCAATCAACAGCGTTTGGTTTAGAGCTTCTGTTCATCTTAATGTTCCCTGAT	65094
QY	869	TTCCAGGAAAGTTTTTTTCCACTATCGCAAACTGAAAAAATATATACATTTACAGACTGGT	928
Db	65093	TTCCAGAGAAATTTTGGCAACTATCGACAGACAGCGCAGCTCTTTCTCATTCGAATCT	65034
QY	929	TTCACTTTTCGTCCTCTCAAAATCGCGTCTGGTTCCTCAATGTCACGCTCCCGGAAGA	988
Db	65033	GGTTCTAGCTATTCATCTAATGCTAACACACTTGTTCCTGTGTGTGATCTCCTCCAGGG	64974
QY	989	ATTTCAATTCCTACAGAATTTTGTTCAAAATATAGTTCTTTGGTACACATGGGATGCATA	1048
Db	64973	TTTAGTTTGGCCTTTTGAGATCTTGTTTAACTTAACATTTGGTGCAGAACGCGTGCTC	64914
QY	1049	CCTGGCCAGCAATTAATATGCTACTTTTTTCCGATTAGTTGATCCCTCGAAGGAGAAATGTG	1108
Db	64913	TCCGGCCAGCTCTTGATCTTGACTTTTATCCGGTTCCTTAATCAGAAAAAATATGATAGA	64854
QY	1109	GCATGCATTGAGCATGCCCTTAGAGAAACTGTAATATAAAGGAGTGCCTTATGATGCC	1168
Db	64853	GCTCTCATAGATCACTGCTTGAAAACTCTTCACCTTGGAGATGTGTGATGAGCCT	64794
QY	1169	GTGAGTGGCTCCTGAGCAGATGATGGGTATCTCAAGGTAGACAACTCCAAAATCT	1228
Db	64793	GCTCAATGGCTTCGGATGAGTACAAGAAAGTGGATATCGAAAGGAAAGCTTCCACTGTCT	64734
QY	1229	CGTCCATCACTTTAGATGATGGGTGGTGATGTAAAGGGTCCCTAGTAACACCATGC	1288
Db	64733	CCTACAATCTCTTGATGATGGACTGTTTATATGATACCGGTTCAGTTTACACCAACA	64674
QY	1289	AAAGTTTATTTTGTGTCAGAGGTTAATGTTTCCATCGGGTCTCCGCAATATTTCT	1348
Db	64673	AGAGTCTATTTCTCTGGACCGAGGTGAATGTTCTTAATCGTGTCTCGACATTACTCC	64614
QY	1349	GAAGACATAGATAACTTTCCTGTTCTTTTGTGATGAGGAGTGGGAGAACTGTAT	1408
Db	64613	AAGTATATCAACAACCTCCCTCAGGCTCTCAITTTGTTGATGAAGATCTTGAAGAGTTCGT	64554
QY	1409	TCTACAGACTTATTACAAAAGCAAGTACTGGAAGTGGTGCAGACAAACATCTATGAG	1468
Db	64553	TCCATGGATCTATCTCCAGCTCTCTTACTCAAAG-----AAGAACCAAGTTATACGAC	64500
QY	1469	AGGATCTTATCAACTCTCGGAAAGGCTTTGTAAATGGTGATAAAAAATTTGAATTTCTT	1528
Db	64499	AGGATATATCTGTTCTTCGGATGGGATGTCTCATGGCGATAAGAAAGTTTGAAGTTCTC	64440
QY	1529	GCATTTTCATCGACCGAGTTCCGGGATAAATTCAGTGTGGATGTTTGCATCAAGACCTGGC	1588
Db	64439	GCCTTTCTCCAGCCAGTTCCGGGGAACCTCTCCTTGGATGTTTGACCAATAGATCGG	64380
QY	1589	CTTACTCGAATGATATAAGAGCTTGGATGGGTGATTTTTCGCAGATCAAGAATGTCGCA	1648
Db	64379	ATTACAGCAGCACATATCAGAGCTTGGATGGGTGATTTTGACCATATAAGAATGTAGCT	64320
QY	1649	AAATATGCTCCAGACTTGGTCAATCTTTTGGTTCCTCCAGAGAGACTTTGAGTGTTCTT	1708
Db	64319	AAATGCTCGAAGGCTTGGTCAATCTTTTAGCTTCTTCAAGAGAGACACTTAATGTTAGG	64260
QY	1709	AGGATGAGATGAAGTATTCCCGATGTAAGAGT-----TCATGGAACCCAGCTATGTC	1762
Db	64259	AGTGATGAGATTGAATGATTCCTGTGATTTGAGATTATATCTTTAGGACACGCTATGTG	64200
QY	1763	TTTTCTGATGGAATTTGTAATAATCTGTGACACTTTGCTCATAGAGTTGCTCCAAAATGT	1822
Db	64199	TTTTCCGATGGAATAGGAAAGATATCAGCTGAATTTCTCTAGAAGGTAGGAGGAAATGT	64140
QY	1823	GGCCTT---CAATATACCCATCTGCTTCCAGATTCGTTATGGTGGATATAAAGTGTT	1879
Db	64139	GGCTTACAGAGTTTCTCCATCTGCTTTCAAATCCGTTATGCCGTTTATAAGGAGTG	64080
QY	1880	GTGGGTGTTGATCCGGATTATCAATGAAGTTGTCTTTTGAAGAGACATGTCCAAATAT	1939

Db	64079	GTGGCTGTTGATCCAAACTCATCAAGAAAACTGTCTCTGAGGAAGAGTATGAGCAAAATTC	64020
QY	1940	GAATCAGACAACATAAAGTTAGATGCTCTTGGATGGAGCAAAATATCAGCCTTGTTATCTTT	1999
Db	64019	GAATCGGAGAACACCAAGCTTGATGTTCTGGCTGGAGCAAGTACCAACCTTGTTATATG	63960
QY	2000	AATCTCAACTGATTACGCTTGTTCTACACTTGGAGTGAAGATGAAGTTCTCGAACAG	2059
Db	63959	AACACACAACGATTACACTTTTGTCTACTCTTGGAGTTACGGACAGTGTGTTGAGAAG	63900
QY	2060	AAGCAAAAGGAAGCTGTAGATCAGCTTCATGCTATCTTCTGATGATCTCTTGAAGCCACAG	2119
Db	63899	AAACAAAGGGAAGTTGTGGATCGCTCGACGAATCTTGAATCATCTTGGAAAGTCTCAC	63840
QY	2120	GAGGCTTTGGAAATTGATGCTCTCTGGAGAGAACACTAATATCTCAAGGCAATGCTAAAC	2179
Db	63839	GAGGCTCTGTGTTTAAATGGCTCCAGGGGAAACACAAAATATCTCAAGGCAATGATCTTG	63780
QY	2180	TGTGTTTATAAGCCGTGACTGAGCCCTTCTTCAATGATGTTGCRAACCTTCGGGCA	2239
Db	63779	TGTGCTATAAACCTGACGCTGAACCTTCTTCAATGATGCTTCAGAAATTCAGGGCA	63720
QY	2240	TCCAAAGTTGCTCGAATTCGCGACTAGATCAAGAATATTTATTCAAAATGGGAAGCAATG	2299
Db	63719	TCTAAGTTGTTGGAACTACGGACAAAAACAAGGATTTTCACTTCTGGTGGAAATCTATG	63660
QY	2300	ATGGGATGTTGGATGAATCCAGAACCTTGGAAATATGGTCAGGTGTTGTTCACTTTACT	2359
Db	63659	ATGGGATGCCCTAGACGAGACCAAGCAACTGAGGATATGGTCAGGTAGTAGTCAGTATCA	63600
QY	2360	GGTGC - TGGACATGGAGAGTTTCTGACGATTTACATCCATTTAATAACAGCAGATCCA	2417
Db	63599	GATCCCATGAGCCGGAAGCGATTCAATCATCACCGGACCTGTTGTTGTCACAAAAC	63540
QY	2418	CCACAGTAATTTCAATCTGAAGGAAATGTTGTTGGCAAAAAATCCATGCTTGCAATC	2477
Db	63539	CCATGCCCTGCATCTCTGTGACGTGCGTGTCTTCAAGCTGTCAATGTGCCAGCTTTAAAT	63480
QY	2478	CTGGTGATATTCGTGTTTTAAAGCTGTAAATGTTGAGCGCTCCACACCATGCT - - - -	2532
Db	63479	CATATGGTGACTGCGTGTGTTGCCGAAAGGGCTTGAGGTAACATTTTCTATGCT	63420
QY	2533	-----AGATTGTTGTATTCCCTCAGAAA-----	2557
Db	63419	TGATTCGTGTTTTTTTGGAGCTTTTATTTGTTTTAAGTCTCGAACTTTTCATTGGAATCTT	63360
QY	2558	-----CGAAAAAGACCTCATCCGAATGATGTTCTGGGAGTATTGGATGGGA	2607
Db	63359	GTGTTCTTTTGTGATAGGCCACCCCAATGAATGTTCTGGGAGTATTAGTAGGAGA	63300
QY	2608	TATCTACTTTTGTCTGGATCAAGACATGATCCCGCCCAAGGCAAGTCCAGCCGATGGA	2667
Db	63299	TATTTACTTTGTATGTTGGATCAAGAATTGGTTCGCCCAAGACGCTCAACCAATGGA	63240
QY	2668	ATATCTCCAGCACCCAGCATACAGTTGGACCATGATGTCACAAATG - - - - - 2714	
Db	63239	CTACACTCTTGAACCACTCAAATCTTGGATCATGATGTCACAATTGAGGTTATAATAA	63180
QY	2715	-----	2714
Db	63179	ATTTGATCACTTTGTTTATTTTTTTGTTTTTTGGCGTTAGGAGATTGATGGCTACCTT	63120
QY	2715	-----AGGAAGTTGAAGAGTACTTCACCAACTATATTGTGAATGACAGATTTGGGAATC	2767
Db	63119	GGTTTCCAGGAAGTTGAAGAGTACTTTGCGAACTACATTTGTGAATGATAGTATTAGGATC	63060
QY	2768	ATAGCAATGCCATGCTGATTATTGCACAGAGAACTGATATGGCCATGAGTGAATCCA	2827
Db	63059	ATCGGAAATGCTCATACCGGCTTTGCGGATGAAGGAACCACTCAAGGGCTTTAGTGACCA	63000
QY	2828	TGCAAAAACCTGCTGAGCTCTTTTCAATTTGCAGTGGACTTTCCAAAGACTGTTGTTTCC	2887
Db	62999	TGCATTGAGGTTGCAAAAGAAGTTTCAACTGCGGTAGATTTCCAAAACATGGTGTGCA	62940

QY 2888 GCTGAATACCATCTCACTGCGCCCTAAAGAAATACCAGACTTCATGGATAAGCCGAC 2947  
Db 62939 GCTGTATACCGACATCTTACGTGAAGAGTATCTCTGATTCATGGAGAACCAGAT 62880  
QY 2948 AAGACCAGCTATATCTCAGAAAGAGTATTGGAAGCTTTTCAGGAAGTGAAGGACAAA 3007  
Db 62879 AAGCCACATATAGTCTGAAGAGCTGATTGGTAAGCTTTTGAAGAGGTAAGAGCGA 62820  
QY 3008 GCACCTCAGGCTAGCTCTATCGGCACCTTCACAAGAGATGTTGCAAGGAGATCATATGAT 3067  
Db 62819 GCTCCACCGTTGATCTGATCAATCAATCACTGATGCTGATGCTCAAGCTTATGAT 62760  
QY 3068 GCTGATATGGAAGTGTATGATGATGGAAGTATACATTCAGCAAGCTTTTGACTACAAAAC 3127  
Db 62759 AAAGACATGGAAGTGTACGATTTGAGGAGTATGTTGATGAAGCTTTCTACCAAGGCG 62700  
QY 3128 GAATATGACACAGCTGGGTAAATTAATGACTACTATGATGATGATGATGATGATGATGAT 3187  
Db 62699 AATTATGATTCAGTTAGTAAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 62640  
QY 3188 ATACTTAGTGTGCATATTAAGGCGCAATCAAACTTTTACCGCGCAAGAAAGATGCTGAG 3247  
Db 62639 ATTCTTAGTGTGATATTAAGGCGCAATCAAACTTTTACCGCGCAAGAAAGATGCTGAG 62580  
QY 3248 GCCATTAGTGTGCTGAGGCGCTTACGAGGAGGCGCAAGGCGCTGCTGCTGCTGCTGCTGCT 3307  
Db 62579 TCGATTGAGAGCGGTTTCGGCGCTTACGAGGAGGCGCAAGGCGCTGCTGCTGCTGCTGCT 62520  
QY 3308 AATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3367  
Db 62519 GAAGAAGAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 62460  
QY 3368 ACATATGAGGCTTCTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3427  
Db 62459 AGTTACTGAGGCTTCTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 62400  
QY 3428 TGTGTTATGACCACTAATCCAGATTAAGAGGACAA 3465  
Db 62399 TGCCTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 62362

RESULT 6  
ATH011977  
LOCUS ATH011977 1538 bp DNA linear PLN 07-JAN-1999  
DEFINITION Arabidopsis thaliana Rdrp gene, partial.  
ACCESSION AJ011977  
VERSION AJ011977.1 GI:4127462  
KEYWORDS Rdrp gene; RNA-directed RNA polymerase.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
REFERENCE 1 (bases 1 to 1538)  
Wassenegger, M.  
Direct Submission  
Submitted (16-OCT-1998) Wassenegger M., Max-Planck-Institut of  
Biochemistry Viroldresearch, Am Klopferspit 18A, 82152  
Martinsried, FRG  
REFERENCE 2 (bases 1 to 1538)  
Schiebel, W., Pelissier, T., Riedel, L., Thalmeir, S., Schiebel, R.,  
Kempe, D., Lottspeich, F., Saenger, H.L. and Wassenegger, M.  
Isolation of an RNA-directed RNA polymerase-specific cDNA clone  
from Tomato  
Unpublished  
FEATURES  
Location/Qualifiers  
1. .1538  
/organism="Arabidopsis thaliana"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/tissue\_type="leaves"  
1. .1538  
gene

CDs  
/gene="Rdrp"  
<1. >1538  
/gene="Rdrp"  
/codon\_start=1  
/evidence=experimental  
/product="RNA-directed RNA polymerase"  
/protein\_id="CAA09894.1"  
/db\_xref="GI:4375829"  
translation="IRAWMGDFDHIRNVAARLGGSSSSRRLTNVRSDEIEVDP  
VELISLGRVYVFDGIGKISAEFAKVKARKGLTEFSPSAQIRVGYGVVAVDPNS  
SKLSLRKSMKFSFSENTKLDYLAWSKYQPCYMMRQLITLLSTLOVDSVPEKKQREV  
VDRDLAITLHELEAHEALGMAPGENTNLRKALILCGYKPDPEFLDSMLQNFRAKSL  
LELRTKTRIFISGGRSMGCLDETLEYGVVYQYSDMPRPFRRFITFVGVVAKNP  
CLHSGDVRVLAQVNPALNHNVDVCFQKGLRPHNECSGSLDGDITFYFCWQELV  
PPRSEPMDTPTETQILDHDTIEVEVEYFANYVNDLSLGIANAHTAFADKPLKA  
FSDPCIELAKKFSTAVDPKTVAAVPOHLYVKEYPDFMEKPKPTVYESKNVIGKLF  
REVERAPPLISIKSFTLDVASKYDKDMVEVDGFEYVDEAFYQKANYDFKLGLMDY  
VGIKT"  
BASE COUNT 442 a 279 c 381 g 436 t  
ORIGIN  
Query Match 20.5%; Score 765; DB 8; Length 1538;  
Best Local Similarity 69.7%; Pred. No. 5,2e-171;  
Matches 1104; Conservative 0; Mismatches 425; Indels 54; Gaps 3;  
QY 1604 ATAAAGAGCTTGGATGGGTGATTTTCGCGAGATCAAGATGTCGCAAAATATGCTGCCAGA 1663  
Db 1 ATACGGCGCTGGATGGGTGATTTTGAACCATATAGAAATAGTAGCTAAATATGCTGCAAGG 60  
QY 1664 CTTGTGCAATCTTTTGGTTCCTCCAGAGAGACTTTGAGTGTTCCTTAGCATGAGATTGAA 1723  
Db 61 CTTGTGCAATCTTTTGGTTCCTCCAGAGAGACTTTGAGTGTTCCTTAGCATGAGATTGAA 120  
QY 1724 GTATTTCCCGATGTAAGGT-----TCATGGAACAGCATGCTGCTTTTCGATGGAATT 1777  
Db 121 GTGATTCCTGATGTTGAGATTATATCTTTAGGGACACGCTATGCTTTTCGATGGAATA 180  
QY 1778 GGTAAATATATCTGGTGACTTTGCTCATAGATTGCTCAAAATGTTGGGCTT---CAATAT 1834  
Db 181 GGAAAGATATACAGTGAATTTGCTAGAAAGGTAGCGAGAAATGTTGCTTACAGAGTTT 240  
QY 1835 ACCCATCTGCTTCCAGATTCGTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1894  
Db 241 TCTCCATCTGCTTTTCAAAATCCGTTATGCGGTTTAAAGGAGTGGTGGCTGTTGATCCA 300  
QY 1895 GATTTCATCAATGAAGTGTCTTTGAGAAAGAGCATGCGAAATATGAATCAACACACATA 1954  
Db 301 AACTCATCAAGAAACTGCTCTGAGGAGAGATGAGCAAAATCGAATCGAGAGACACC 360  
QY 1955 AAGTTAGATGCTTGGATGGAGCAAAATATCAGCCTTGTATCTTAATCGTCAACTGATT 2014  
Db 361 AAGCTTGATGTTTGGCGTGGAGCAAGTACCAACCTTGTATATGACAGACAACTGATT 420  
QY 2015 AGCCTCTTGTCTACACTTGGAGTGAAGATGAAGTCTCGAACAGACAAAGAAAGAGCT 2074  
Db 421 ACACCTTGTCTACTCTTGGAGTTACGACAGCTGTTTGGAGAGAAACAAGGAAAGTT 480  
QY 2075 GTAGATACGTTGATGCTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2134  
Db 481 GTGGATCGCTCGACGCAATCTTGACTCATCTCTTTGGAAGCTCAGCAGGCTCTTGGTTTA 540  
QY 2135 ATGTCTCTGGAGAGAACACTAATATCTCAAGCAATGCTAAACTGTTGTTAAGCTT 2194  
Db 541 ATGGCTCAGGGGAAACACAAATATCTCAAGGCAATGATCTTGTGGGCTATTAAGCTT 600  
QY 2195 GATGCTGAGCCCTTTCTTTCAATGATGTTGCAAACTTCCCGCATCCCAAGTTGCTCGAT 2254  
Db 601 GAGCGTGAACCTTTCTTTCAATGATGCTTCAGAAATTTTCAGGCAATCAAGTTGTTGAA 660  
QY 2255 TTGCGGATAGATCAAGAAATATTTATCCAAATGGAAGAACAAATGATGGGATGTTGGAT 2314  
Db 661 CTACGGACAAAACAAAGGATTTTCTGTTGGAAGATCTATGATGGGATGCTGCTAGAC 720

QY	2315	GAATCCAGAACCTTGGAAATATGGTCAGGTGTTTGTTCAGTTTACTGGTCTGGAGCATGGA	2374
DB	721	GAGACAGAACCGTGGAGTATGGTCAGGTAGTGCAGTATTCAGATCC-----	770
QY	2375	GAGTTTCTGACGATTTCATCCATTTAATACACGATCCACCAACAGTAATTTCA	2434
DB	771	-----CATGAGCCGGGAAGCGGATTCATC	795
QY	2435	CTGAAGGAAATGTGTTTGCACAAAAATCCATGCTTGCATCTGGTGATATTCGTGTT	2494
DB	796	ATCACCGGACCTGTTGTTGTCACAAAAACCATCCCTGCATTTCTGGTACGTCGCGT	855
QY	2495	TTAAGGCTGTAATGTTTCGAGCGCTGCACACATGGTAGATTGTTGTTGATTCCCTCAG	2554
DB	856	CTTCAAGCTGTAATGTCACCATCTTAAATCATATGGTGGAGCTGCGTTGTGTTCCGCCAA	915
QY	2555	AAAGGAAAAGACCTCATCCGAATGAATGTTCTGGAGTGATTTGGATGGGATATCTAC	2614
DB	916	AAGGGCTTGAGGCCACACCCAAATGAATGTTCTGGAGTGATTAGATGGAGATATTAC	975
QY	2615	TTTGTGTTGGGATCAAGACATGATCCCGCAAGTCCACGCGATGGAATATCCT	2674
DB	976	TTTGTATGTTGGATCAAGAAATGTTCTCGCCAAAGACGCTGAACCAATGGACTACACT	1035
QY	2675	CCAGCACCCAGCATACAGTTGGACCATGATGTCACAAATGAGGAAGTTGAGAGTACTTC	2734
DB	1036	CCTGAACCAACTCAAACTTTGGATCATGATGTCACAAATGAGGAAGTTGAGAGTACTTT	1095
QY	2735	ACCAACTATATGTTGAATGACAGTTTGGGAATCATAGCAAAATGCCATGCTGATTGCA	2794
DB	1096	CGAACTACATTTGTAATGATAGTTTAGGATCATCGCGAATGCTCATACGCGCTTTGCG	1155
QY	2795	GACAGAGAACCTGATATGCCATGATGATCCATGCAAAAAAAGTTGCTGAGCTCTTTTCA	2854
DB	1156	GATGAAGAACCACTCAAGCGCTTATGACCCATGCATTTGAGCTTGCAAAAGAAGTTTCA	1215
QY	2855	ATTGCTAGTGGACTTCCAAAGACTGGTGTCCGCTGAAATACCATTCTAGTTGCGCGCT	2914
DB	1216	ACTGCGGTAGACTTCCAAAAAAGTGGTGTGCTGCTGTATACCGCAACATCTTAGCTG	1275
QY	2915	AAAGATACCCAGACTTCATGGATTAAGCGGACAGACCAAGCTATATCTCAGAAAGAGTT	2974
DB	1276	AAAGAGTATCCTGATTTCATGTGAGAAACCGGATAGCCAAATATGATGTCGAGAAGCTG	1335
QY	2975	ATTGGAAGCTTTTCAGGAAGTGAAGCAAGACCTCAGGCTAGCTCTATCGCGACC	3034
DB	1336	ATTGGTTAAGCTTTTAGAGAGGTAAAGAGCGAGCTCCACCGTTGATCTCGATCAATCA	1395
QY	3035	TTCAAGAAGATGTTGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGGAATTTGAA	3094
DB	1396	TTCAACGCTTGATGTGGCCTCAAAGTCTTATGATAAGACATGGAAGTTGACGATTTGAG	1455
QY	3095	GATTACATTGACGAAGCTTTTCACATCAAAAACTGAATATGACAAACAGCTGGGTAA	3154
DB	1456	GAGTATGTTGATGAAGCTTTCTACCAAGAGGGGAATATGATTTCAAGTTAGGTAATTG	1515
QY	3155	ATGGACTACTATGGCATAAAAAC	3177
DB	1516	ATGGACTACTATGGCATAAAAAC	1538

RESULT	7
AP004143	
LOCUS	80554 bp DNA linear HTG 21-MAR-2002
DEFINITION	Oryza sativa (japonica cultivar-group) chromosome 2 clone CJ1699_D12, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION	AP004143
VERSION	AP004143.1 GI:15451464
KEYWORDS	HTG; HTGS_PHASE2.
SOURCE	Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA, clone:OJ1699_D12.
ORGANISM	Oryza sativa (japonica cultivar-group) Eukaryota; Viridiplantae; Streptophyta; Embryophyta Tracheophyta;

REFERENCE AUTHORS TITLE	JOURNAL REFERENCE AUTHORS TITLE JOURNAL
1. H. J. ... 2. ... 3. ... 4. ... 5. ... 6. ... 7. ... 8. ... 9. ... 10. ... 11. ... 12. ... 13. ... 14. ... 15. ... 16. ... 17. ... 18. ... 19. ... 20. ... 21. ... 22. ... 23. ... 24. ... 25. ... 26. ... 27. ... 28. ... 29. ... 30. ... 31. ... 32. ... 33. ... 34. ... 35. ... 36. ... 37. ... 38. ... 39. ... 40. ... 41. ... 42. ... 43. ... 44. ... 45. ... 46. ... 47. ... 48. ... 49. ... 50. ... 51. ... 52. ... 53. ... 54. ... 55. ... 56. ... 57. ... 58. ... 59. ... 60. ... 61. ... 62. ... 63. ... 64. ... 65. ... 66. ... 67. ... 68. ... 69. ... 70. ... 71. ... 72. ... 73. ... 74. ... 75. ... 76. ... 77. ... 78. ... 79. ... 80. ... 81. ... 82. ... 83. ... 84. ... 85. ... 86. ... 87. ... 88. ... 89. ... 90. ... 91. ... 92. ... 93. ... 94. ... 95. ... 96. ... 97. ... 98. ... 99. ... 100. ...	1. ... 2. ... 3. ... 4. ... 5. ... 6. ... 7. ... 8. ... 9. ... 10. ... 11. ... 12. ... 13. ... 14. ... 15. ... 16. ... 17. ... 18. ... 19. ... 20. ... 21. ... 22. ... 23. ... 24. ... 25. ... 26. ... 27. ... 28. ... 29. ... 30. ... 31. ... 32. ... 33. ... 34. ... 35. ... 36. ... 37. ... 38. ... 39. ... 40. ... 41. ... 42. ... 43. ... 44. ... 45. ... 46. ... 47. ... 48. ... 49. ... 50. ... 51. ... 52. ... 53. ... 54. ... 55. ... 56. ... 57. ... 58. ... 59. ... 60. ... 61. ... 62. ... 63. ... 64. ... 65. ... 66. ... 67. ... 68. ... 69. ... 70. ... 71. ... 72. ... 73. ... 74. ... 75. ... 76. ... 77. ... 78. ... 79. ... 80. ... 81. ... 82. ... 83. ... 84. ... 85. ... 86. ... 87. ... 88. ... 89. ... 90. ... 91. ... 92. ... 93. ... 94. ... 95. ... 96. ... 97. ... 98. ... 99. ... 100. ...

**COMMENT**

FEATURES source

BASE COUNT	23035 a	16948 c	17802 g	22769 t
ORIGIN				

Query Match 14.8%; Score 551.4; DB 2; Length 80554;  
Best Local Similarity 60.1%; Pred. No. 2e-120;  
Matches ll08: Conservative 0; Mismatches 646; Indels 89; Gaps 8;

QY	734	GAAACTCTGTTATAGCTTCCTTTAAGGAAACTCTGATGATCAGTGGGTGAGGACACA	793
Db	8359	GAAGATCCATATTCAACTATTTTGGGATCATACAGATGATCAATGGACGACGACACG	8418
QY	794	GATTTCCCTCCATCTT---GGATAGGGTACTCTTCTAGCTTATGTTTSCAGTCCGTAGG	850
Db	8419	GATTTTACTTCATCTCTAGCATGGGCAATCATATATTTTATGCTTGAGTACCAGT	8478
QY	851	GGTGTTCGTCTTCCAAATTTTCGAGGAAAGTTTTTCCACTATCGAGAACGTTGAAACAAT	910
Db	8479	CGATGTGATCTCCCAACATTCGTGACTACTCTTTTACTATCATGAGTACAATCATGAC	8538
QY	911	ATTACTTTACAGACGGTTTCACCTTTTTTCGTCCTCAAAATCGGCTCTCGTGTCCCAAT	970
Db	8539	TTGAGTCCGGGAGTGG---GGGGTATCCATATTCAAGTACTCCGCTTTGTTCGGATT	8595
QY	971	GTCCAGCCTCCGGAAAGAAATTTCAATTCCTCACAAGATTTTGTTCAAAATTAGTCTCTTG	1030
Db	8596	GTCAATCTCGTGG-----TTATGTCCCTATGAGATACTCTTCAAGATCAACCAATTG	8649
QY	1031	GTACAGCATGGATGATACCTGTGGGCCAGATTAATGTCTACTTTTTTCCGATTAGTTGAT	1090
Db	8650	GTTCAGAAATGGACACACTTAGCGGACCGACAGTTGATGACAGCTTCTTTCCGTCTGGTCAGC	8709
QY	1091	CTCGAAGGAGAAATGTGCGATGCATTCAGCATGCGTTAGAGAACTGTACTATATAAG	1150
Db	8710	CCAGCTTTTGTACTATTGACCATATAAAGCGTCGACTTGAATGATGTCTATTTGAAA	8769
QY	1151	GAGTGTGTTATGATCCCGTGGGTGCTCACTGAGCAGTATGATGGGTATCTCAAGGGT	1210
Db	8770	AAGACTGCTGTAATCCCAACAGTTGGCTATCTGAACAATACTCAAAATTCGGAGATCA	8829

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.

1 Sasaki,T., Matsumoto,T. and Yamamoto,K.  
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC  
clone:OJ1699\_D12  
2 Published only in Database (2001)  
2 (bases 1 to 80554)  
Sasaki,T., Matsumoto,T. and Yamamoto,K.  
Direct Submission  
Submitted (05-SEP-2001) Takuji Sasaki, National Institute of  
Agricultural Sciences, Rice Genome Research Program; Kamondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail:tsasaki@nias.affrc.go.jp, URL:<http://rgp.dna.affrc.go.jp/>,  
Tel:81-298-38-7441, Fax:81-298-38-7468)  
The nucleotide sequence of this BAC clone was generated by  
combining Monsanto and RGP-Japan sequencing data.  
NOTE: It currently consists of 1 contigs. Gaps between the contigs  
are represented as runs of N. The order of the pieces is believed  
to be correct as given, however the sizes of the gaps between them  
are based on estimates that have provided by the submitter. This  
sequence will be replaced by the finished sequence as soon as it is  
available and the accession number will be preserved.  
\* NOTE: this is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

Location/Qualifiers
1. .80554

1. .80554

```
1. 80334
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="2"
```

```

/cione="001699_DL2"
a 16948 c 17802 q 22769 t

```

Db	9884	AAAGCAGCAATGATGGGATGCTTGGATGAACACACGTACGCTGAAGTAGTGGGCAAGTG	9943
QY	2345	TTTGTTCAGTTTACTGGTCTGGACATGGAGAGTTTCTTGACGATTTACATCCATTTAAT	2404
Db	9944	TTCAATTCAGACTACTCGGGTGAATG	9971
QY	2405	AACAGCAGATCCACCAACAGTAATTTTCATCTGAAGGAAATGTGGTTGCGCAAAAT	2464
Db	9972	-----ACAAATGACAGGTTCAACCTGACAGGAAAAGTTGTGATTGCCAAAAC	10018
QY	2465	CCATGCTGCATCCTGGTGATTTCTGGTTTAAAGCTGTAAATGTTTCGAGCGCTGCAC	2524
Db	10019	CCTTGCTCCACCCAGGATATACGGAATTCCTCATGCTGTGATGTTCTCTTTTGAC	10078
QY	2525	CACATGTAGATTTGGTTGTTATCCCTCAGAAAGGAAAAAGAC	2567
Db	10079	CACATGTTTAACTGTGCTGCTTCTCCACAGCAAGCAAGGC	10121
RESULT 8			
AP004880			
LOCUS			
DEFINITION			
Oryza sativa (japonica cultivar-group) chromosome 2 clone P0487D09, 162558 bp DNA linear HTG 23-MAR-2002			
*** SEQUENCING IN PROGRESS ***, in ordered pieces.			
ACCESSION			
AP004880			
VERSION			
AP004880.1 GI:19698300			
KEYWORDS			
HTG; HTGS-PHASE2.			
SOURCE			
Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA, clone:P0487D09.			
ORGANISM			
Oryza sativa (japonica cultivar-group)			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.			
REFERENCE			
1			
Sasaki, T., Matsumoto, T. and Yamamoto, K.			
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, PAC			
clone:P0487D09			
JOURNAL			
Published Only in Database (2002)			
REFERENCE			
2 (bases 1 to 162558)			
Sasaki, T., Matsumoto, T. and Yamamoto, K.			
Direct Submission			
JOURNAL			
Submitted (20-MAR-2002) Takuji Sasaki, National Institute of			
Agrobiological Sciences, Rice Genome Research Program; Kannondai			
2-1-2, Tsukuba, Ibaraki 305-8602, Japan			
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,			
Tel:81-298-38-7441, Fax:81-298-38-7468)			
COMMENT			
NOTE: It currently consists of 1 contigs. Gaps between the contigs			
are represented as runs of N. The order of the pieces is believed			
to be correct as given, however the sizes of the gaps between them			
are based on estimates that have provided by the submitter. This			
sequence will be replaced by the finished sequence as soon as it is			
available and the accession number will be preserved.			
* NOTE: This is a 'working draft' sequence.			
* This sequence will be replaced			
* by the finished sequence as soon as it is available and			
* the accession number will be preserved.			
FEATURES			
source			
1. 162558			
Location/Qualifiers			
/organism="Oryza sativa (japonica cultivar-group)"			
/cultivar="Nipponbare"			
/db_xref="taxon:39947"			
/chromosome="2"			
/clone="P0487D09"			
BASE COUNT			
44809 a 36152 c 35022 g 45360 t 215 others			
ORIGIN			
Query Match			
14.8%; Score 551.4; DB 2; Length 162558;			
Best Local Similarity 60.1%; Pred. No. 1.8e-120;			
Matches 1108; Conservative 0; Mismatches 546; Indels 89; Gaps			



JOURNAL  
COMMENT

Submitted (24-JUL-1998) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University,  
St. Louis, MO 63108, USA  
e-mail: rwilson@watson.wustl.edu

MAPPING: Clones were assigned to the YAC map by hybridization by  
M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted  
by M. Marra, WashU, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we provide a small overlap between  
neighboring submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate  
chemistry; an attempt was made to resolve all sequencing problems,  
such as compressions and repeats; all regions were covered by  
sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 3' clone is F8M12, 200 bp overlap. Actual start of this clone  
is at base position 1 of F2P3; actual end is at 23552 of F8M12.

NOTES:

Coding sequences below are predicted from computer analysis, using  
the program Genefinder (P. Green and L. Hillier, ms in preparation).

FEATURES  
source

1. 74406  
/organism="Arabidopsis thaliana"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="IV"  
/map="unknown"  
/clone="F2P3"  
complement(979. .1536)  
/gene="F2P3.6"  
complement(979. .1536)  
/gene="F2P3.6"  
/note="contains similarity to Pisum sativum disease  
resistance response protein 206-d (GB:U11716)"  
/codon\_start=1  
evidence=not\_experimental  
/protein\_id="AAC35542.1"  
/db\_xref="GI:3600055"  
/translation="MTNQIVQVFSFLSVLLQSSVSVPKSFDLKPKKHFLVYL  
HN1AYDGNANATATIVKPLGHDHSGFGLIIINPVTLQNYLSKPVARAQGYF  
YNNKTYNANVATLVFNSTKHKGTETIMDNFGLQPLQADLSIVGTGDFLMTGRIA  
TFKTLTQGSKYCEVMNKLKYC"  
complement(2210. .2293)  
/product="tRNA-Tyr"  
/note="codon recognized: UAC"  
complement(2779. .3078)  
/gene="F2P3.7"  
complement(2779. .3078)  
/gene="F2P3.7"  
/note="similar to translation initiation factor IF-1"  
/codon\_start=1  
evidence=not\_experimental  
/protein\_id="AAC35543.1"  
/db\_xref="GI:3600056"  
/translation="MIRCORASGGRGANRSKPAKPOVKEGSKNTVIEGLVTESLPNG  
MFVYDLENGNIIIGYCGTKRKIRFIILPGDKVKVMSYDSTKGLIIFRMSRD"  
complement(4455. .10458)

gene

CDS

/gene="F2P3.8"  
complement(join(4455. .5858,6701. .7012,7300. .8407,  
9995. .10458))  
/gene="F2P3.8"  
/note="similar to several Arabidopsis thaliana disease  
resistance proteins"  
/codon\_start=1  
evidence=not\_experimental  
/protein\_id="AAC35544.1"  
/db\_xref="GI:3600057"  
/translation="MASSSSNRWDYDFSPRGEDVRNRFSLHLKESEKGIQVTFRD  
DIKRSHTIGHELRAAIRESKISVYVFNSENASSWCDELTEIMKCEQGLKMPY  
FKYVDPDIRKOTGKFGFLETCGCKTEERQHNWRALTDAAINLGDHPQWDEAY  
KTTISKVLEKLNATPSRFDNDLVGMEAHAKMESLLECSQGVIRIVGPGAGVGK  
TIARALYNVHENENISTEMENVRESYGEAGLDYGLKLHQREFLSKLLDQDLRV  
RILGAIERLKSQKVIILDDVDNIEQLAKAKENQWGNKSRIYVITQNLQSLVSHD  
INMTQVAPSKQEALTIFCQAFKQSSDDLLKLALEFTEITLAGHLPAUKLVGSFM  
RKGKEWEFSLPTLSRLDGEVEKVLKVGNDLHDEKDLFLHACIFSGQHENYLR  
QMIIANNDTYVSFGLQVADKSLIQKFGDIEMHSLRLQGLKGVYRKOSIYEPGKR  
FLMAKETGVSNNGTGTGVLIGSLDMCEIKELYISEKTFEEMRNLYVLFYMSPP  
IDDKMKVLQPEGLSYLPQRLHWDAYPLEPFPSPRECELVLPSSIKNLOHLIL  
SGVQPLRLMTNLNSSLNLEITLNLMEATKLNRLDLCWCELVLPSSIKNLOHLIL  
LEMSCKKLEIIPTNINPSLEYLHFRTCTRLQPEPEISTNIRLLNLIGTAVRPPS  
VKYSKIDEIMERAKVRKVHPVLEKLEKLENKELETPRILYKLPQADMIDISY  
CINIISLPKGSVSALTAVNCEISQILHGFNRKSIHLNFINCLKUGRAQEKIHR  
VIHQSYIADYLPGEHVPAYFSYRSTSIHNSKVDLSKFNRFKCLVGLGAKRF  
EGCDIKFYKQFCKREVYVVKHLDSPLLKSHLCMCELFMPHPPTWELLPEH  
LEVSPESRGLLYKCEKCGQLQFLEPHSTSEFRLSPHLYLGGSNIGNSSSIEIIF  
VDOESSSDSEELIYADQESSSGTIEIHAEREGTNRKRKSVMRWIKVGAARKMGLSLE  
CLKPWTR"  
complement(14859. .17002)  
/gene="F2P3.9"  
complement(join(14859. .15293,15428. .15574,15656. .16209,  
16261. .16801,16868. .17002))  
/gene="F2P3.9"  
/note="similar to initiation factor IF2-beta (Pfam:  
GTP\_BFYU.hmm, score: 226.67)"  
/codon\_start=1  
evidence=not\_experimental  
/protein\_id="AAC35536.1"  
/db\_xref="GI:3600049"  
/translation="MTLLELSKRTGESVAVLQSLINVGTFSSFEFTISVDVAELLA  
MEIGINVRROHSTGESEILPRPPVTVMGVDHDKTSLDLALRNTSVARAGGJTOH  
VGAFVYGMPSDSTSTFIDTPGHAFASEMRARGAATDVIVVAADQVMPQLEAI  
AHARSANPVVAINKDKPGANPEKYKIITSEIGIELEDIGGNVQVQKDNFGWGLIN  
FLQFGLAKSTGLDKLEBALILQAVMDLKAQVDPQAGYVVEARLDKRGRLATIV  
KAGTLVRGQHVVGQWGLRAIRDMIGKTTDRATPAMPVEIEGLKLPJMGAGDDVIV  
ESEERARMLSEGRKRYKDRLLKAEAEARLEAEARLESEEGFVRYELFVYKSDY  
QGTAAVADALRTLNSPOVSNIVHSGVGAISHDLDAQACGACGIVFNKGGSTGN  
LSAAGSVKVPFHRVYIHLEDIGNLIVKEKAPGVSEMEVSGEAEVLSIFKILGRRT  
EDGVNAGCKVMDGRVCBSGLMLRLSRGSEVFECSASLKRKQDQVEQVKGNEGGL  
FGDWNDFRGVDVQCMPEVIRKPKFISSEAVRIEC"  
complement(19889. .21494)  
/gene="F2P3.10"  
complement(join(19889. .19924,20035. .20134,20239. .20435,  
20635. .20791,21177. .21277,21393. .21494))  
/gene="F2P3.10"  
/note="similar to vacuolar ATPases; coded for by A.  
thaliana cDNA T43489; coded for by A. thaliana cDNA  
T04528; coded for by A. thaliana cDNA R64882; coded for by  
A. thaliana cDNA T46025; coded for by A. thaliana cDNA  
AA042676; coded for by A. thaliana cDNA T21569; coded for  
by A. thaliana cDNA T45327; coded for by A. thaliana cDNA  
AA042335; coded for by A. thaliana cDNA W43660"  
/codon\_start=1  
/protein\_id="AAC35545.1"  
/db\_xref="GI:3600058"  
/translation="MNDGVSQRQQQVVRFRTRQAEKANEISVSAEEFNIEKLQVL  
EAEKKIRQDYKKEKQADVRKIDYSQMLNARIKVLAQDDIVANMKDQAADLLN  
VSRDYAYKQLLKDLIVCLLKEPSVLLRCREBGLDVEAVLDDAEEYAGAKVH  
APEVAVDNKFILPPPPKSNDPHGLHCSGGVVLASRDGKIVCENTILDARLDVAFRMKLP  
VIRKSLFGQVTA"  
25791. .26654

gene

CDS

gene

CDS

gene









QY	1802	CATAGAGTTGCCTCAAAATGTGGCCTTCAATATACCCCATCTGCTTCCAGAGATTCGTTAT	1861
Db	54251	AACAAAGTTGCACAGAAGTGTGGATGAGTCATGTCCCTCTCGCTTTCAAATTCGATAC	54310
QY	1862	GGTGGATATAAAGGTGTTGTGGGTGTGTATCCGGATTCATCAATCAAGTGTCTTTGAGA	1921
Db	54311	GGTGGCTACAAAGGTGTGATGTGTGTACCCGACGTTCCTCCGAAAGTGTCTCTCGGT	54370
QY	1922	AAGNCATGTCGAAATATGAATCAGACAACATAAAGTTAGATGTCCTTGGATGGAGCAAA	1981
Db	54371	GATAGTAGTCTTAATTTGACTCGAACACACAGATGCTCAAGCTTACCAGGTGGACAG	54430
QY	1982	TATCAGCCTTGTATTCTTAATCGTCAACTGATTACGCTCTTCTACACTTGGAGTGA	2041
Db	54431	TCGATGCTTGTCTTCTTAACCGGGAGATCATTTGCTTTTCGACACCTTTGGAAATGAA	54490
QY	2042	GATGAAGTTCGACACAGAGCAAGGAGCTGTAGATCAGCTTGATCTATCTTGCAT	2101
Db	54491	GATGCAATGTTTGAAGCGATGCAAGCGGTGCATTTATCTATGCTTGGGNAATGCTTGA	54550
QY	2102	GATTCCTTTGAAGCAGCAGGAGGCTTTGGAAATTTGATGCTCTCTCGGAGAGCACTAATTT	2161
Db	54551	GACCGCATCGACACTTAATGTTCTGCAGAAATTCAGCGGAGAAATTCCAAGAAATTTG	54610
QY	2162	CTCAAGCGAATGCTAAACTGTGTTATAGCCTGATGCTGAGCCCTTCTTTCATATGATG	2221
Db	54611	CTAGTTAAGATGTGCTTCAAGGATATGACCGAGTTCAGAACCTTACCTCTCAATGATG	54670
QY	2222	TTGCAAACTTCGCGCATCCAAAGTTGCTCGATTTCCGGACTAGATCAAGAATATTTAT	2281
Db	54671	CTTCGTGTGCACACGAGAGCCAGCTTCTGAACTAAAGAGCAGATGCAGGATACCTGTA	54730
QY	2282	CCAAATGGAAGAACATATGATGGGATGTTGGATGATTCACAGAACCTTGGAAATGTCAG	2341
Db	54731	CCGAAAGCAGGATCTTGTATGCTGGTTCGATGGATGAAATGGTATCTCGAGATGCGCAA	54790
QY	2342	GTGTTGTTGATGTTACTGGTGTGGCATGAGAGTTTCTGACCATTTACATCAATTT	2401
Db	54791	GTGTATGTTCTGTGTAACCTGACTAAAGCGGAACTGAAATCTCGCATCAGAGCTACTTT	54850
QY	2402	AATAACAGCAGATCCCAACAGTAATTCATCTCGAAGGAAATGTGTTGTCGAAA	2461
Db	54851	-----CGCAAGATGATGAGGAACATCTGTGTTATTGGAAAGTGTGTCGAGCAAA	54904
QY	2462	AATCCATGCTTCATCGTGGATATTCGTTTAAAGGCTGTAAATG	2510
Db	54905	AACCAATGCTTCACCTGGAGATTTAGAGATTCCTTGATGCTATATG	54953
RESULT 11			
ATCHRIV31		198301 bp DNA linear	PLN 16-MAR-20
LOCUS	ATCHRIV31	198301 bp DNA	linear
DEFINITION	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 31.		
ACCESSION	AL161531		
VERSION	AL161531.2	GI:7267806	
KEYWORDS			
SOURCE	Arabidopsis thaliana.		
ORGANISM	Arabidopsis thaliana		
	Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.		
REFERENCE	1 (bases 1 to 81065)		
AUTHORS	Pohl, T., Weizenecker, T., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.		
JOURNAL	Unpublished		
REFERENCE	2 (bases 73007 to 171954)		
AUTHORS	Peters, S.A., van Staveren, M., Dirkse, W., Stiekema, W., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.		
JOURNAL	Unpublished		
REFERENCE	3 (bases 75956 to 76343)		
AUTHORS	Volckaert, G., Grymonprez, B., Voet, M., Robben, J., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.		
JOURNAL	Unpublished		

```

REFERENCE
AUTHORS      4 (bases 152454 to 156565; 171236 to 172629)
              Robben,J., Grymonprez,B., Voickaert,G., Mewes,H.W., Lemcke,K. and
              Mayer,K.F.X.
JOURNAL      Unpublished
REFERENCE    5 (bases 1 to 198301)
AUTHORS      EU Arabidopsis sequencing project.
TITLE        Direct Submission
JOURNAL      Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
              Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
              lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project
              Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
              Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
              E-mail: michael.bevan@bbsrc.ac.uk
COMMENT      Information on performance of analysis and a more detailed
              annotation of this entry and other sequences of chromosomes 3, 4
              and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/chal/
              this fragment has an overlap with ARCHIV30 at the 5' end and an
              overlap with ARCHIV32 at the 3' end.
              Location/Qualifiers
FEATURES     source
              1..198301
                /organism="Arabidopsis thaliana"
                /variety="Columbia"
                /db_xref="taxon:3702"
                /chromosome="4"
                /number="5"
                /gene="AT4g11080"
                complement(join(5135..5242,5338..5422,5499..5590,
                5675..5845,5935..6000,6090..6195,6284..6690,7204..7509))
                /gene="AT4g11080"
                complement(join(5135..5242,5338..5422,5499..5590,
                5675..5845,5935..6000,6090..6195,6284..6690,7204..7509))
                /note="strong similarity to 98b - Daucus carota (carrot),
                PID:d1033793"
                /codon_start=1
                /product="98b like protein"
                /protein_id="CAB81209.1"
                /db_xref="GI:7267807"
                /translation="MSTVSSDPAHAKKSRKALKOKNELVESPVDKQKTKSFE
                KDLMEQANLEKKKIEKEDTEDLLKDEILRKKEVEQKLTLEKLLKQKMKPEFQNM
                TFAFSQSLAQTEEEKKQKAKKQCAETKPSPIYLWCKDNWNEKVKQKPEADFKETS
                NILGAKWGSIAEEKKPYEEKYQADKEAYLQVITKEKREAMKLLDDQKQKTAHEL
                LDQYLHVPQAEHNDKKKAKKIDPLKPKQPSAYLIYANERRAALKGNKGVIVAK
                MAGEEWKPQKPYQDQAKKKEIYLOEMEGYKRTKEEAMSKQKEEFEMKLHK
                QEALQLLKKEKTDNIITKTKETAKNKKNNVDNPKPKPTSSVPLFCCKARKSVLE
                EHPGNNSTVTAHTSLSKWMLGEEQVYNSKAAELMEAYKKEVEYNKTKTSS"
                complement(5135..5242)
                /gene="AT4g11080"
                /number=1
                complement(5243..5337)
                /gene="AT4g11080"
                /number=1
                complement(5338..5422)
                /gene="AT4g11080"
                /number=2
                complement(5423..5498)
                /gene="AT4g11080"
                /number=2
                complement(5499..5590)
                /gene="AT4g11080"
                /number=3
                complement(5591..5674)
                /gene="AT4g11080"
                /number=3
                complement(5675..5845)
                /gene="AT4g11080"
                /number=4
                complement(5846..5934)
                /gene="AT4g11080"
                /number=4
                complement(5935..6000)
                /gene="AT4g11080"
                /number=5
                complement(6001..6089)
                /number=5
                complement(6090..6195)
                /gene="AT4g11080"
                /number=6
                complement(6196..6283)
                /gene="AT4g11080"
                /number=6
                complement(6284..6690)
                /gene="AT4g11080"
                /number=7
                complement(6691..7203)
                /gene="AT4g11080"
                /number=7
                complement(7204..7509)
                /gene="AT4g11080"
                /number=8
                8882..10458
                /gene="AT4g11090"
                complement(join(8882..9670,9765..10057,10242..10458))
                /gene="AT4g11090"
                complement(join(8882..9670,9765..10057,10242..10458))
                /gene="AT4g11090"
                /note="similarity to other hypothetical proteins -
                Arabidopsis thaliana
                contains EST gb:AI995110.1, N38106, R90644"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB81210.1"
                /db_xref="GI:7267808"
                /translation="MKLWESISNLQNNYLIKVAATLITCLAFEFVFRGQSPV
                QSVTGNSSQISFTSVLSDNEDQIPVDIEVEKCDLFTGKWKIKDPLGFIYNESCGI
                VDAHQNCITNGRPDSGFLNWKPKNDCLSPRDSLRFLQMRNKSALIGSIAIARNH
                VESLLCMLSTVEKPEYVHDENYSKRHFPSYNTVSNISWPFLLVQADIFEDSNGVS
                SAAQLHLKLDINTDLPFLDYAIISSEGENFLKTAHYENANPVGCHGCPESNMV
                DLGFDYATSLRHVMDFIAKSKTKGMIFRTSIPDHFEDGEWHNGTCKKTPYVEE
                AVENKNTKILRDVEINOFERVVTEMGQESENKLLDFAFMLITRDGHPGYPYREFRP
                FDKDKNVNDCLHWCPLPGPIDLNDVILEIIVNGRTCK"
                complement(8882..9670)
                /gene="AT4g11090"
                /number=1
                complement(9671..9764)
                /gene="AT4g11090"
                /number=1
                complement(9765..10057)
                /gene="AT4g11090"
                /number=2
                complement(10058..10241)
                /gene="AT4g11090"
                /number=2
                complement(10242..10458)
                /gene="AT4g11090"
                /number=3
                11978..12608
                13297..14213
                /gene="AT4g11100"
                join(13297..13861,13903..14213)
                /gene="AT4g11100"
                /note="EST GB:T43920 marks 5'end, EST GB:AA395296 marks
                3'end
                contains EST gb:AA395296, AI997464.1, T43920"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB81211.1"
                /db_xref="GI:7267809"
                /translation="MTLEHEILKENHETLLKDYESLQERIKKAEAYEAIKLHENKA
                KELEVSNKRLLEECMKERKAKVKTFFEMKMTSERTAIVDELSKQNELLGKK
                KEELVKMENKVELAEKEDVYKCEAYLKSLYDAEVAVSQSAVTSVSGVETDNL
                QGQGENANQNNNAVGTGIHRLCTQVLFNAIMISDDNNPPIESNISHSRSGDRNQ
                EEEGRNRSATENRASGETFTVPSANHPPTFPSPSSSSSSSSSSSSSSSSSSSSSSSS
                EWATPKGPKGDSNKP"
                13297..13861

```

```

/gene="AT4g11100"
/number=1
13862..13902
/gene="AT4g11100"
/number=1
13903..14213
/gene="AT4g11100"
/number=2
16403..20912
/gene="AT4g11110"
join(16403..16347,16590..17213,17271..17866,18203..18321,
18736..19568,19624..19859,19922..20239,20334..20393,
20721..20912)
/gene="AT4g11110"
/note="EST GB:AA404849 spans an intron in 3' nontranslated
region and marks 3' end
similarity to photomorphogenesis repressor COP1 -
Arabidopsis thaliana, F1R2:R01112
Contains Prokaryotic membrane lipoprotein lipid attachment
site AA405-415;Trp-Asp (WD-40) repeats signature
AA799-813;Trp-Asp (WD-40) repeats signature AA919-933
contains EST gb:AI99586.1, T46420, AA404849"
/codon_start=1
/product="COP1 like protein"
/protein_id="CAB81212.1"
/db_xref="GI:7267810"
/translation="MDEGSGVDYSRIDEADVAHLQKNSQSPKPNIEVREYKQV
OREADGNKGHVELIGNPCSSRNQNDGVDVEELTVKCEGSSMAIYGRPSRAR
LEMRSQFLHRELDLPGSSSMKKVIDRGTVSILRNAGKMSLPETSNQALIAV
NGEANHDTWERNPVVEALSHGKIKMLSGSFQFFVRKTLKGKGVTFRPPNN
RSKARNMDQTVASSGSALVIANISIKSIPLAAYDDTHRGCGEGLSRLWLKSE
RQVNAECMIPIFQIVDHDVCSHQGVFLCDLRPSFKIFKFNAYKIVYVSGQRSEF
DSNNKETLSQLENPLVRRRLGDTSSLSIPAKKOKSGSPSSRQMPFQAGGVNIOTE
Query Match      8.4%; Score 314.6; DB 8; Length 198301;
Best Local Similarity 55.2%; Pred. No. 3.6e-64;
Matches 589; Conservative 0; Mismatches 539; Indels 21; Gaps 3;

QY 1277 GTAACACATGAAAGTTTATTTTGTGTCAGAGGTTAATGTTCCAAATCGGGTCTC 1336
DB 26178 GTTACACCTCAAGACTATCTGTGTGGTCCAGAGCTTGAGACTGCAAAATGTTGTG 26237
QY 1337 CGCAATATTCTGAAGACATAGATAACTTCTCGTGTCTTTTGTGATGAGAGTGG 1396
DB 26238 AGAAGTTGACAGCACTGCTCGGATTTCATGAGATTACTTTTGTGGAAGAGATTGG 26297
QY 1397 GAGAACTGTATTCACAGACTTATACCAAGCAAGTACTGGAAGTGGTCTAGG--- 1453
DB 26298 AGCAAGCTTCCCGCAATGCTCTCTGTGACTCCAAAGGAGGCTATTTTGTGAAGCCC 26357
QY 1454 -----ACAAACATCTATGAGAGGACTTATCAACTCTCGGAAAGGCTTTGTAATTGGT 1507
DB 26358 TGTAGAACCAACTTATACAGAGGTGTGTCATCTTGGAGAGGATTAACGTTGGG 26417
QY 1508 GATAAAAAATTTGAATTTTTCATTTTCATCGAGCCAGTTGCGGATTAATTCAGTGGG 1567
DB 26418 CTTAGAGGTTTGTAGTCTCTGTGCTTTTTCAGGAGCCAACTCGAGGAAATTCGCTCG 26477
QY 1568 ATGTTTGCATCAAGACCTTACTGCAATGATATAAGAGCTTGGATGGTGATTTT 1627
DB 26478 ATGTTTGTCTTACAGAGAAATAAAGCGGAAGATATAAGAGATGGATGGTGTGTTT 26537
QY 1628 TCGCAGATCAAGATGTCGCAAAATGTCGCGAGACTTGGTCAATCTTTTGGTTCCTCC 1687
DB 26538 CGTAAATCCGAGCACTTCCAAATGCTGCTAGAGTGGGTCACTTGTTCAGTCTTCG 26597
QY 1688 AGAGAGCTTTGAGTGTCTTTCAGGATGAGATTGAAGTTATTCGCGATGAAGGT---- 1743
DB 26598 CGGCAACACTTATTTCTCGGTGCACAGAGATGTGGAGCAGATCCCGACATCGAAGTGACA 26657
QY 1744 --TCATGGACACGCTATGCTTTCTGATGGAATTTGGTAAATATCTGGTCACTTTGCT 1801
DB 26658 ACTGATGGGTGCTATTACTGCTTCTCGGATGGCAATGGGAAATTTCACTTGCATTGCC 26717

```

```

QY 1802 CATAGAGTTGCCCTCAAAATGTCCTTCAATATACCCCATCTGCTTTCCAGATTCGTTAT 1861
DB 26718 AACCAAGTTGCCAGAGTGTGGATTGAGTCAITGTCCTTCTGCTTTCAAAATCGATAC 26777
QY 1862 GTTGATATAAAGGTGTGTGGGTGTGTGATCCGGATTCATCAATGAAGTTGCTTTGAGA 1921
DB 26778 GTTGGCTACAAAGTGTGATTTGTGTGACCGGAGTTCCTTCGGAAGAGTGTCTCTCGCT 26837
QY 1922 AAGAGCATGTCGAAATGATGAATCAGACACATAAAGTTAGATGCTCTTGGATGGAGCAAA 1981
DB 26838 GATAGTATGCTTAAATTTGACTCGACACACAGAGATGCTGAACGTTACCAAGTGGACAG 26897
QY 1982 TATCAGCTTGTATCTTAATCTGCTCAACTGATPACGCTCTTGTCTACACTTGGAGTGAA 2041
DB 26898 TCGATGCTTGTCTTCTTAAACCGGAGATCATTTGCTTGTGTCGACCTTGGAAATAGAA 26957
QY 2042 GATCAAGTCTTCCACACAGAGCAAAAGGAGCTGTAGATCAGCTTGTGATGCTATCTTGCAT 2101
DB 26958 GATGCAATGTTTGAAGGATGCAAGGGTGCAATTTATCATGCTTGGATATGCTTGAA 27017
QY 2102 GATTTCTTTGAAGCACAGGAGGCTTTTGAATGATGCTCTCTCGGAGAACACTAATATT 2161
DB 27018 GACCGGATGCGACACTTAATGTTCTGCAGAAATTTGACGAGAGAAATTCCAAGAAATTG 27077
QY 2162 CTCAAGCAATGCTAACTGTGTTTAAAGCTGATGCTGAGCCCTTCTTTCAATGATG 2221
DB 27078 CTAGTTAAGATGCTGCTTCAAGGATATGACCGAGTTTCAAGCTTACCTCTCAATGATG 27137
QY 2222 TTGCAAAACCTTCCGGCGCATCCAAAGTTTCTCGATTTCGGGACTAGATCAAGAAATTTATT 2281
DB 27138 CTTCGTGTGCACACAGAGAGCCAGCTTCTGAACTAAAGAGCAGATCGAGGACTATTGA 27197
QY 2282 CCAATGGAAGACATGATGGATGTTTGGATGAATCCAGACCTTGGATATGTCAG 2341
DB 27198 CCGAAAGAGGAGTCTTTGATCGGTGATGATGAAATGGTATPCCTGGAGTAGGCCAA 27257
QY 2342 GTGTTTGTTCAGTTTACTGCTGCTGGACATGAGAGGTTTCTGACGATTTTACATCCATTT 2401
DB 27258 GTGATGTTCTGTAAACCTGACTTAAGCGGAAGTGAATCTCGCATCAGAGCTACTTT 27317
QY 2402 AATAACAGCAGATCCCAACACAGTAATTTCAATCTTGAAGGAAATGTTGTTGTCAGAAA 2461
DB 27318 -----CGCAAGATGTATGAGGAACATCTGTTTATTTGGAAGGTGTCGTCGACGAAA 27371
QY 2462 AATCCATGCTTGCATCTCTGTTGATATTCGTTTATTTAAAGGCTGTAAATG 2510
DB 27372 AACCAATGCTTCCCTCGAGATATTAGATTCTTGTATGCTATATG 27420

```

RESULT 12  
OSJN00104 121492 bp DNA linear PLN 17-JUL-2002  
LOCUS OSJN00104  
DEFINITION Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJNB0048502,  
complete sequence.

ACCESSION AL606653  
VERSION AL606653.2 GI:21741220  
KEYWORDS HTG.

SOURCE Oryza sativa.

ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE  
1  
AUTHORS Han,B., Feng,Q., Huang,Y.C., Li,Y., Zhu,J.J., Zhao,Q., Hu,X.,  
Liu,Y.L., Mu,J., Yu,Z., Chen,L., Fan,D.L., Weng,Q.J., Zhang,L.,  
Lu,Y.Q., Yu,S.L., Liu,X.H., Lu,T.T., Zhang,Y.J., Lu,Y., Li.C.,  
Li,T., Zhang,Y., Hu,H., Jia,P.X., Qian,Y.M., Ying,K., Zhou,B.,  
Chen,Z.H., Hao,P., Zhang,L., Wu,M., Zhang,R.Q., Guan,J.P., Fu,G.,  
Wang,S.Y., Ren,S.X., Lv,G., Lin,W., Gu.W.Q., Zhu,G.F., Tu,Y.F.,  
Jia,J., Yin,H.F., Zhang,Y., Cai,Z., Chen,J., Kang,H., Chen,X.Y.,  
Shao,C.Y., Sun,Y., Hu,Q.P., Zhang,X.L., Zhang,W., Wang,D.J., and  
Ding,C.W., Sheng,H.H., Gu,J.L., Chen,S.T., Ni,L., Zhu,F.H. and

TITLE Direct Submission  
 JOURNAL Submitted (27-JUN-2002) Han Bin, National Center for Gene Research,  
 Chinese Academy of Sciences, 500# Cao Bao Road, Shanghai 200233,  
 CHINA. E-mail enquiries: bhanncgr.ac.cn. Clone requests:  
 bhanncgr.ac.cn  
 REMARK Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC  
 clone: OSJNB0048E02.  
 COMMENT On Jul 12, 2002 this sequence version replaced gi:15594110.  
 Web site: <http://www.ncgr.ac.cn>  
 ----- Summary Statistics  
 Assembly program: Phrap

-----  
 NOTE: This is a complete sequence.  
 Genes were identified by a combination of several methods: Gene  
 prediction programs including Fgenesh (<http://www.softberry.com/>),  
 gensean (<http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkESM  
 (<http://genemark.biology.gatech.edu/GeneMark/>), tRNAscan-SE (Sean  
 Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>), searches of the  
 complete sequence against NCBI none redundant protein database (nr)  
 (<ftp://ncbi.nlm.nih.gov/blast/db>) and the EST database at NCGR.  
 Location/Qualifiers  
 1. 121492  
 /organism="Oryza sativa"  
 /variety="Nipponbare"  
 /sub\_species="japonica"  
 /db\_xref="taxon:4530"  
 /chromosome="4"  
 /clone="OSJNB0048E02"  
 /clone\_lib="CUGI-OSJNB"  
 BASE COUNT 34036 a 25389 c 25204 g 36813 t 50 others  
 ORIGIN

Query Match 8.4%; Score 314.2; DB 8; Length 121492;  
 Best Local Similarity 52.8%; Pred. No. 4.7e-64;  
 Matches 865; Conservative 0; Mismatches 733; Indels 41; Gaps 7;  
 QY 953 TCGGCTCTGGTTCCTCAAGTCCAGCTCCGGAAGAAATTCCTCAAGATTTG 1012  
 DB 99416 TCCATGTGGTCCCTTGTCATGTCGAGGAGACTATTCAGTCCCTATGAAGTCTG 99475  
 QY 1013 TTCAAATAGTCTTTGGTACAGATGGATGCATACCTGGCGGAGCAATTAATGTCTAC 1072  
 DB 99476 TTTCGTCTGAATCTCTCATGCATATGGGGAAGATATCGCCAAAGCATGTTAATGCCGAT 99535  
 QY 1073 TTTTCCGATAGTATCTCCGAGGAGAAATGTGGCATGCATGAGCATGCCCTTAGAG 1132  
 DB 99536 CTGTTTAAAGCCCTTCAAGAACTACCGGTTGATGTCCTCAG-----GAGAAATTTTGA 99589  
 QY 1133 AAACGTACTATATAAGAGAGTCTGTTATGATCCGCTGAGTGGCTCCTACAGCAG--- 1189  
 DB 99590 AAATGCAAAATAAAGTCTACCTGCTACGGGCTTTACAATTTATCCAAACAGAGGCT 99649  
 QY 1190 --TATGATGGGTATCTCAAGGGTAGACAACTCCAAATCTCCGTCCTCATCTTAGATG 1247  
 DB 99650 TATAGCATGAAGAGAGGCAATTAATTTTGTCTCCCAATGAAGGTGAAGGAGGTGAAA 99709  
 QY 1248 ATGGGTTGGTGTATGAAGAAGGGTCTTAGTAACACACCATGCAAAAGTTTATTTTGTGTC 1307  
 DB 99710 GAAATGATGAAGTGTATAGAGTTACATAACTCCCTCCAAATATTCGTGTTGAC 99769  
 QY 1308 CAGAGGTAAGTTTCCAAATCGGGTCTCCGCAATTAATTTGAGACATAGATAACTTTC 1367  
 DB 99770 CTGAAGAAGAGGTTACAAATATATGTTGTTAAGCATCATCTGTCATATGCTTCTGACATTG 99829  
 QY 1368 TTCGTGTTCTTTCTGATGAGGAGTGGGAGAACTGTAATCTA-----CAGACT 1418  
 DB 99830 TGAGAGTATCTCTCTGATGAGATTCGAGTAAGCTTTCTCTAATGCAATCTCAGCTA 99889  
 QY 1419 TATTACCAAAAGCAAGTACTGGAAGTGGTGTGAGACAAACATCTATGAGAGGATCTTAT 1478  
 DB 99890 GAATTGAACAAGGTTCTCTTCTAAGCCCTTTCAAACCTGCTATATATTATCGGATTTCTCT 99949

QY 1479 CAACTCTGGGAAAGGCTTTGTAATTGGTGATAAAAAAATTTGAATTTCTTGCATTTTCAT 1538  
 DB 99950 CCATTTAAAAAAGATTTTCCATTTGGTCCCAAGAACTTTGAATTTTGGCTTCTCAG 100009  
 QY 1539 CGAGCCAGTTGCGGATATAATTCAGTGTGGATGTTTGCATCAAGACCTGSCCTTACTGCAA 1598  
 DB 100010 CAACTCAACTCGGGGAAATCTGTTTGGATGTTTCTTATATGCTTCTTGAATGCTG 100069  
 QY 1599 ATCATATAAGAGCTTGGATGGGTGATTTTTCGAGATCAAGAATGTGCGAAAATATGCTG 1658  
 DB 100070 GAGTATAGAAGATGGATGGGCCACTTTGAGAAATCCGTTCAAGTTCTAAATGTGCAG 100129  
 QY 1659 CCAGACTTGGTCAATCTTTGGTTCCTCCAGAGAGACTTTGAGTGTCTTAGCATGAGA 1718  
 DB 100130 CTAGAATGGGTCAAGCTTTCAGCTCTCCAGGCAAACTTTGAAGTCTTCAATGGGATG 100189  
 QY 1719 TTCAAGTATTTCCTCGATGAAGGT-----TCATGGAACCAAGCTATGTTCTTCTCGATG 1772  
 DB 100190 TGAAGTATTCCAGATATTGAGATCACACATGATGGTCCCAATACATATTTCTGATG 100249  
 QY 1773 GAATGGTAAATATCTGGTGAATTTGCTCATAGATTTGCTCAAAATGTGGCTTCAAT 1832  
 DB 100250 GTATTGGGAAGATATCTTTGAGATTTTCCCAAGCGAGTTGCACACCACGTTGGATTAGATC 100309  
 QY 1833 ATAC-----CCCATCTGCTTCCAGATTCGTTATGGTGGATATAAGAGTGTGTGGGTG 1886  
 DB 100310 CTACTAACCTTCTTCCAGCTTTCAATAAGGTATGGTGGTACAAAGAGTCAATTTGCCA 100369  
 QY 1887 TTGATCCGGATTCATCAATGAAGTGTCTTTGAGAAGAGCAATGTCGAATATGAATCAG 1946  
 DB 100370 TTGATCCTATGCTCTATTGATCTTTCTGCTGCTCTAGTATGAAGAAATTTTGAAGTCTG 100429  
 QY 1947 ACAACATAAGTTAGATGCTTTGGATGGGCAAAATATCAGCCTTGTATCTTAAATCGTC 2006  
 DB 100430 AGAGCAGATGTTGAACATTTACAAGTTGGAGCAAGTCTCAGCCATGCTATGTAATCGTG 100489  
 QY 2007 AACTGATTAGCTCTTGTCTACACTTTGGAGTGAAGATGAAGTCTCGAACAGAGCAAA 2066  
 DB 100490 AAATTTATTTCTCTCTCAACATTTGGTATAAGGATGAATTTTGTGCAATGCAAC 100549  
 QY 2067 AGNAGCTGTAGATCAGTTGATGCTATCTGTCATGATCTTTTGAAGGCACAGAGGCTT 2126  
 DB 100550 AAGATGAGATGCGTGAACAGAGAA---TGTACCAACAAAGAGTGTCTTTGTCAG 100606  
 QY 2127 TGAATTTGATGCTCTCTGGAGAGAACTAATATTTCTCAAGGCAATGCTAAATCTGGTT 2186  
 DB 100607 TCCTAGGAAACTTGGTGGTTCTGAAACAAAGACAGCAGTGAAGATGTTGCTGCAAGGT 100656  
 QY 2187 ATAGCCTGATGCTGAGCCCTTTCTTCAATGATGTTGCAACCTTCCGCGCATCCAAGT 2246  
 DB 100667 ATGAGCCTAGTTTCAGAGCCTTACCTGTCAATGATTTTAAAGGCCCATCAGGAAATAGGC 100726  
 QY 2247 TGCTCGATTTGCGGACTTAGATCAAGAAATATTTTCCAAATGGAAGAACAAATGATGGAT 2306  
 DB 100727 TTACTGACATTAGAAGTATGATGTAATACATGTTCCAAAGGCCGCTGTTCTTATTGGT 100786  
 QY 2307 GTTTGGATGAATCCAGAACCTTGGATATGGTCAGTGTGTTTTCAGTTTACTGTTGCTG 2366  
 DB 100787 GTTTGGATGAACAGGTGTATTAGAATATGTCAGGTTTACATTCGAATTACAAAGAAC 100846  
 QY 2367 GACATGAGAGTTTCTGACGATTTACATCCATTTTAAACACAGATCCACCAACAGTA 2426  
 DB 100847 GCAAGAAC-----AAAGACAGTAATCAATCATATTTTATATGATGACGGGAAA 100900  
 QY 2427 ATTTCAATTCGAGGGAATGTTGTTGCAAAAAATCCATGCTTCGATCTCTGTTGATA 2486  
 DB 100901 CAGCCACAGTTGTTGGAAGAGTTGCAATCAAAAAATCGTCTCCATCTCTGTTGATA 100960  
 QY 2487 TTGCTGTTTAAAGGCTGTAAATGTTTCGAGCGCTGCACACATGGTAGATTGTTGTTAT 2546  
 DB 100961 TCCGAGTACTTGAAGCCATCTATGACCCTGATTGTTGGGCAATGTTGATGTTCTGTTCT 101020  
 QY 2547 TCCTCAGAAAGGAAAAAG 2565

```
||||| 101021 TCCCTCAAGAGGAGAAAG 101039
Db 101021 TCCCTCAAGAGGAGAAAG 101039

RESULT 13
AX204840
LOCUS AX204840 3591 bp DNA linear PAT 30-AUG-2001
DEFINITION Sequence 2 from Patent WO0155407.
ACCESSION AX204840
VERSION AX204840.1 GI:15394184
KEYWORDS thale cress.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 3591)
Beclin,C., Elmayan,T., Mourrain,P. and Vaucheret,H.
Novel sg2 plant gene and use thereof
Patent: WO 0155407-A 2 02-AUG-2001;
Rhubio (FR) : INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (FR)
FEATURES
source
1..3591
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
1..3591
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAC60174.1"
/db_xref="GI:15394185"
/translation="MGSEGNMKSVTVQVSGFGESTTAKQLDYLEDEGVIVWRCR
LKTSWTPPGSYNPFADSNIPDSIDYKKVEPHAFVFEAFESGARMADAQGNLI
LDQPLKVSIGPKNPYLNQRRITVYKLAGITLEIGTLVSRDQFFVSWRAEGVDFL
VDFPNTCKFCPRKFAFKDAVMHAINCDYKLELLVRIQTVQRYKTLHGVLIL
QLASSPRVWRTADDIDYDTPGDLLDDDDPWIRTDFTQVGAIGRCHSYRLVLSPLY
ENKLRALDYFMRVQERVWRPPIRNEPCGEPEVSDHFECHHKEGISLEIMLV
SNVLHGVNQFQTLERFEDLNRNQKDNVNIASLKLHCYIKRPVEDAYKRLKIQEWI
QKPKLLGSHQESDSEIRLRLITRAYCLUPPELVEVNSRNLRLRYKAVRFLRVTF
DMESQITINSVLVFAVPIVKDLSFSQKTYFVKRYSILTDGFKLCGRKYSFLA
FSANQLDRSAMPFADGKT RVSDIKTMGKFKDNVAKARMGLCFSTVATVDVM
PHEVDTEPIERNGVSDIGITIPDLADEVMEKLDVHYSPCAQIYRAGVPGV
VAVPWSKSDGIRALRDSMKFKSKHTILEIGSWTFQPGFLNRQITLLSLVGPDE
IFWDMQESMLKLNRLDIDTDAVEFLVATSCAEQGNATMILSAGPKTEPHLRGML
SVRIQAQLWGLRKSRIFVTSGRWLMGLDEAGILEHGOCFTQVSKPSTENECSEKHS
SFRETIDLEVVAGYVAIAKNPCLHPGDFVLEDAVDVQPLHMHMIDCLIFPQGRPHI
NEASGLDQICNAHVDQSKLIPNKRSPAMHYDAEESKLSGRVNHQDIIIDFARN
LANEQLTICNAHVHADRSYEGAMDECLLAEALATADVDFPKTKIYVMPFHLKP
LYPDFMGKEDYQYKKNKILGRLYRRVKEVYDDEASSEESTDPSAIPYDAVLEIPG
FEDLIPFANGHKLVDGOLIGLLGOYKVKOEIEIVTGHWSMPKYSKKOGELKERLK
HSYNSLKKERKVFETIPDHENLSEENKILYKKASAWYHYTHPEVKKSLLEQD
PDESSHAMLSFANWIAADYLRKIKIRSENGSIDSAKPVDSLAKFLAQL"
BASE COUNT 1022 a 714 c 858 g 997 t
ORIGIN

Query Match 8.2%; Score 305.8; DB 6; Length 3591;
Best Local Similarity 51.0%; Pred. No. 7e-62;
Matches 1013; Conservative 0; Mismatches 917; Indels 57; Gaps 10;

Qy 1263 TAAGAAGGTCCTAGTACACCATGCAAGTTTATTTTGGTCCAGAGGTTAAGTTT 1322
Db 1232 TCAGAAGACTAGTAATACCCCAACAGAGCCATTGCTTACCCCAAGAGTTGAGCT 1291
Qy 1323 CCAATCGGTTCTCGGAATATTCTGAAGACATAGATACTTCTTCTGTTCTTTT 1382
Db 1292 CCACAGGGTACTCAGGAGATACAAGCTGTGCTGAAGATTTTTCGGGGTAACTTCA 1351
Qy 1383 TTGATGAGGAGTGGGAGAACTGTATTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1429
Db 1352 TGGATGAAAGTATGACACCAATAATTCGAATGTTCTCTCTCTCTCTCTCTCTCT 1411
Qy 1430 GCAAGTACTGGAAGTGGTGT-----CAGGACAAACATCTATGAGGAGTCTTAT 1478
Db 1412 TGAAGGATTGCATCAAGTTCCTCTCCCAAGAGACTAGCTTTTAAAGAGTGAAGA 1471
```







**ORGANISM**  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Brassicales; Brassicaceae; Arabidopsis.  
**REFERENCE**  
 1 (bases 1 to 4013)  
 Mourrain,P., Beclin,C., Elmayan,T., Feuerbach,F., Godon,C.,  
 Morel,J.B., Jouette,D., Lacombe,A.M., Nikic,S., Picault,N.,  
 Remoue,K., Sanial,M., Vo,T.A. and Vaucheret,H.  
**TITLE**  
 Arabidopsis SGS2 and SGS3 genes are required for  
 posttranscriptional gene silencing and natural virus resistance  
 Cell 101 (5), 533-542 (2000)  
**JOURNAL**  
 MEDLINE  
 PUBMED  
**REFERENCE**  
 2 (bases 1 to 4013)  
 Beclin,C., Mourrain,P., Vaucheret,H. and Elmayan,T.  
**AUTHORS**  
 Direct Submission  
**TITLE**  
 Submitted (28-FEB-2000) Biologie Cellulaire, INRA, Route de  
 Saint-Cyr, versailles 78026, France  
**FEATURES**  
 Location/Qualifiers  
 1..4013  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="III"  
 join(<1..2715,3138..>4013)  
 /product="SGS2"  
 /note(1..2715,3138..4013)  
 /note="RNA-dependent RNA polymerase-like protein"  
 /codon\_start=1  
 /product="SGS2"  
 /protein\_id="AAF73959.1"  
 /db\_xref="GI:8164028"  
 /translation="MGSGNMKKSIVTQVSGIGFGEISTAKQTLDTYLEDEVGIVWRCR  
 LKTSVPGSYNFEIATSNRTIPKAGILETGLVSRDDDFVSRAGVDFL  
 LDQQLPYSGLPNKSYNQRRRTIPKAGILETGLVSRDDDFVSRAGVDFL  
 VDFPNTCKFCETATSFKDAMVHAINCQDKLELLARDIDQIVRGYKTLHGFVLL  
 QLASGPVRMYRTADDIDYTVGDLDDDDPWIRITFTQVAGLQVRYKTLHGFVLL  
 ENKLRTALDYFMKRVQERVRWPRIRNEPFCGEVSDHFFCIHHKEGISFELMFLV  
 NSVLHGVNFQMLQTERFDLLRNQPKDNVIANSLHCLTKYKVPVEDAYKRLKLQVEM  
 QXNKLGSHEQSEDSIRIRVITPTRAYCLPPELSNRLRYKVAARFLRVFT  
 MEESQRTINSVLSFVAPVYKDLTSSFSOKYVEKRVKSILTDGFLGCTGYSTFADVM  
 FSAQMDIRNSAFEDUKTRVSDIKLWNGKFKDKNVAKCAARMGLCFSTYATVDVM  
 PHEVETSDPIERNQYVFSDDGIGTITLPDLAEVMEKLLKLVHSPQAYOIRVAGFQV  
 VARWPSQDGIKRLARDSGMGSFKHTILEICSWTPQPGFLNQRLITLISVLGVQPD  
 IFWDMSEMLYKRLIDDTDAVFVLTASCAEQGNTAAIMLSAGFKAPTEPHRGLMGL  
 SSVRIAQMLKREKSRIFVTSGRLWMLGCLDEAGILEHQCFIQVSKSINCSFKHGS  
 REKETDLEVVGYVATAKNPLCPGDPVIRLEADVQLHHMDTLFPQKGDPRPT  
 NEASGDLGDIAYANVDOKLPPNRKSPYAMHWDAAEKSLRGAHVHQDIIDFPAARN  
 LANEQDITGCIHVAHVRSEGADEICLLLAATAVDFFPKTGIVSMPPHLKPK  
 LYPDFMGKEDIQYKSNLIGLRVRYKVEYDEGAASSESTDPSPALPYDAVLEIKF  
 FLDPIPEANGHCLYDQGLIGLQYKVKQEEIVGTGHINSMPKYTSKKOGELKERLK  
 HSYNSLKKEFRFVETPDHLENSEENKILYEKKAASVHTVTHPEWYKSLSELDQ  
 PDSSHAAMLSFAWTAAADYLIARIKIRSEMGSIDSAKPVDSLAKFLAQRLL"  
 1119 a 778 c 948 g 1168 t  
**BASE COUNT**  
 1119 a 778 c 948 g 1168 t  
**ORIGIN**  
 Query Match 5.9%; Score 220.4; DB 8; Length 4013;  
 Best Local Similarity 51.5%; Pred. No. 1.4e-41;  
 Matches 733; Conservative 0; Mismatches 641; Indels 48; Gaps 8;  
 1263 TAAGAAGGCTCCTAGTAACACCATGCAAGCTTATTATTTTGTGTCACAGGTAATGTTT 1322  
 1232 TCAGNAGACTAGTAATTAACCCACACAGACCTATTGCGCTACCCCCAGAGGTGAGCTCT 1291

1292 CCAACAGGGTACTCAGGAGATACAAAGCTGTGCTGAAGATTTTTCGGGTAACTTTCA 1351  
Ddb  
1383 TTGATGAGGAGTGGGAGAACTGTATTC-----TACAGACTATATACCAAAA 1429  
Qy  
1352 TGGATGAAGATGCAGACCATAAATCGAATGTTCTCTTACTTTGTTGCTCCATTG 1411  
Ddb  
1430 GCAGTACTGGAAGTGGTGT-----CAGGACAAAACATCTATGAGAGATCTTAT 1478  
Qy

Db 1412 TGAAGGATTTGACATCAAGTCTTCTCCAGAGACCTACGTTTTTAAAGAGTGAAGA 1471  
QY 1479 CAACCTCGCGAAGAGGCTTTGTAATTTGGTGATATAAAATTTGAATTTCTTGCATTTTCAT 1538  
Db 1472 GCATATTAAACCGATGGGTTTAAACTATGTGTAGAAAAACAGTTTTCTAGCATTTCTCAG 1531  
QY 1539 CGAGCCAGTTGCGGATTAATTCAGTGTGGATGTTTCATCAAGACCTGGCCTTACTGCRA 1598  
Db 1532 CCAATCAACTGAGAGACCGCTCTGATGGTCTTTGCTGAAGACGGGAAACACGTTGT 1591  
QY 1599 ATGATATAAGAGCTTGGATGGGTGATTTTTCGCAGATCAAGAAATGTCGCAAAATATGCTG 1658  
Db 1592 CAGATATAAACAATGATGGGAAGTT---CAAAGACAAGATGTGGCAAAATGTGCTG 1648  
QY 1659 CCAGACTGTGGTCAATCTTTTGGTTCCTCCAGAGAGACTTTGAGTGTCTTTAGSCATGAGA 1718  
Db 1649 CTAGATGGGCGCTGTGCTTCTCCACATATGCCACTAGATGTCATGCCCTCACGAGG 1708  
QY 1719 TTGAAGTTATTCGGATGTAAAGGTTTCATGGAACCACTATGCTTTCTGATGGAATTG 1778  
Db 1709 TTGA---CACTGAGTTCCAGATATTGAGAGAAATGGGTATGTTTCTCTGACGGAATTG 1765  
QY 1779 GTAAAAATATCTGGTATTTGCTCATAGATGGCTCAAAA-----TGTGGCCTTCAAT 1832  
Db 1766 GTACAATCACACCTGACCTCGCTGACGAAGTAAATGGAGAAACTTAAGTTGGATGTGCAC 1825  
QY 1833 ATACCCCATCTGCTTCCAGATTCGTTATGATTAAGGTGTTGTGGTGTTCATC 1892  
Db 1826 ACAGCCCTGTGCTTATCAGATAGCTTACGAGGTTTCAAGGGGTGTTGCTCGTGGC 1885  
QY 1893 C-----GGATTCATCAATGAAGTTGCTTTGAGAAAGACATGTCGAAATATGAATCAG 1946  
Db 1886 CATCAAAAGTGATGGAATCAGGCTAGCCCTTCGAGACAGTATGAAGAAGTTCTTTTCCA 1945  
QY 1947 ACAACATTAAGTTAGATGCTTTGGATGGAGCAAAATATCAGCCTTGTATCTTAATCGTC 2006  
Db 1946 AACATACGATCTTGAGATCTGTTCCCTGGAGGAGGTTTCAACCTGGGTCTTAAATCGGC 2005  
QY 2007 AACTGATTACGCTCTTGTCTACACTTGGAGTGAAGATGAAGTTCTCGAACAGAGCAAA 2066  
Db 2006 AGATAATTAACCTTCTATCCCTACTAGTGTCCGGATGAATATCTGGGATATGCAGG 2065  
QY 2067 AGAAGCTGTAGATCAGCTTATGCTATCTTTCATGATGTTTGAAGGCACAGGAGCTT 2126  
Db 2066 AATCCATGCTCTATAAATGTAACCGCATCTTGTATGATACAGATGTGTCATTTGAAGTTC 2125  
QY 2127 TGGATTGATGCTCTCGAGAGAACACTAATATTCTCAAGGCAATGCTAAACTGTGTT 2186  
Db 2126 TCACGGCATCATGTGCTGAACAGGAAACACTGCAGCTAT---CATGCTTAGTCAGGTT 2182  
QY 2187 ATAGCCCTGATGCTGAGCCCTTTCTTCAATGATGTGCAAAACCTTCGGGCATCCAAAGT 2246  
Db 2183 TCAAACCAAAACACCGAGCCGATCTACCGGGATGTTGCTTCACTCAGTCAAGATTCACAAC 2242  
QY 2247 TGCTCGATTTGGGACTAGATCAAGATATTATTCAAATGGAAGAACATGATGGAT 2306  
Db 2243 TCTGGGGTCTCAGAAAAATCTCGTATTTTGTACTTTCAGGAGGTGGCTAATGGTT 2302  
QY 2307 GTTTGGATGAATCCAGAACCTTGAATATGTCAGGTGTTTGTTCAGTPTTACTGTGCTG 2366  
Db 2303 GCCTAGACGACGAGGATACTTGAACATGGCAATGCTTTATTCAAGTCTCTAAACCGT 2362  
QY 2367 GACATGGAGAGTTTCTGACGATTTTACATCCATTTAATAACAGCAGATCCACCACAGTA 2426  
Db 2363 CTATAGAAAATTTGTTCTCCAAACATGGGTCTCGTTTTTAAAGAGACAAAGACAGATCTGG 2422  
QY 2427 ATTTTCATCTGAAGGAAATGTTGTTGTCAAAAAATCCATGCTTGCATCTGCTGTGATA 2486  
Db 2423 AAGTAGTT---AAGGCTATGTAGCCATGCTAAGATCCCTGTCTTCAACCAGGGATG 2479  
QY 2487 TTCGTGTTTAAAGCTGTAAATGTTCGAGCCCTGCACACATGTGATGTTGTAT 2546

Db 2480 TAAGGATTTTAGAAGCTGTTGATGTACCCAGCTGCATCACATGTATGACTGCCCTTATTT 2539  
QY 2547 TCCCTCAGNAAGAAAAGACCTCATCCGAATGAATGTTCTGGGAGTGTATTGGATGGG 2606  
Db 2540 TCCCTCAGAAAGGTATAGGCCCGCATACAAACGAGCTTCTGGCAGTGACCTTGACGGG 2599  
QY 2607 ATATCTACTTTTGTGTTGCTGGGATCAAGACATGATCCCGCAA 2648  
Db 2600 ACCTGTACTTTTGTGGCTTGGGATCAGAAACTCATCCCTCCCA 2641

Search completed: November 5, 2002, 22:34:30  
Job time : 11044 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 19:16:42 ; Search time 142 Seconds  
(without alignments)  
8057.824 Million cell updates/sec

Title: US-09-782-874-1  
Perfect score: 3731  
Sequence: 1 GAATATCTTCTTACTT.....AGTTTCATCTTCTTCTTAA 3731

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3731	100.0	3731	4	US-08-811-583-1
2	57.2	1.5	906	1	US-08-100-874-1
3	53.4	1.4	7218	1	US-08-232-463-14
4	40	1.1	162450	1	US-08-345-882-1
5	36.8	1.0	2277	1	US-08-676-967-2
6	36.8	1.0	2277	1	US-08-676-974-2
7	36.8	1.0	2277	2	US-08-098-487-2
8	36.6	1.0	7218	1	US-08-232-463-14
9	36.4	1.0	3191	4	US-08-936-165A-194
10	35.4	0.9	17341	4	US-08-415-946-1
11	35	0.9	1248	4	US-09-134-001C-342
12	35	0.9	1275	2	US-08-911-445-4
13	35	0.9	1275	4	US-08-182-983-4
14	35	0.9	1458	2	US-08-911-445-19
15	35	0.9	1458	4	US-08-182-983-19
16	35	0.9	1796	2	US-08-911-445-17
17	35	0.9	1796	4	US-08-182-983-17
18	35	0.9	1845	2	US-08-911-445-5
19	35	0.9	1845	4	US-08-182-983-5
20	35	0.9	2046	2	US-08-911-445-6
21	35	0.9	2046	4	US-08-182-983-6
22	35	0.9	5613	2	US-08-463-418-1
23	35	0.9	111282	4	US-08-754-250-3
24	34.8	0.9	1028	4	US-08-858-207A-222
25	34.8	0.9	6002	4	US-08-345-882-4
26	34.8	0.9	19446	4	US-08-961-527-51
27	34.6	0.9	1883	4	US-08-991-677-3

## ALIGNMENTS

## RESULT 1

US-08-811-583-1  
; Sequence 1, Application US/08811583  
; Patent No. 6218142  
; GENERAL INFORMATION:  
; APPLICANT: Wassenecker, Michael  
; APPLICANT: Riedel, Leonhard  
; APPLICANT: Schiebel, Winfried  
; APPLICANT: Sanger, Heinz  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN  
; RNA-DIRECTED RNA POLYMERASE (RDRP)  
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RDRP)  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & NEAVE  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/811,583  
FILING DATE: 05-MAR-1997  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley, James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: MPG-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-596-9000  
TELEFAX: 212-596-9090  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3731 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
ORIGINAL SOURCE:  
ORGANISM: Tomato  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 194..3535  
US-08-811-583-1

Sequence 86, Appl  
Sequence 62, Appl  
Sequence 395, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 21, Appl  
Sequence 3, Appl  
Sequence 1660, Ap  
Sequence 59, Appl  
Sequence 1, Appl  
Sequence 73, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 26, Appl  
Sequence 1983, Ap

28 34.6 0.9 19390 4 US-08-961-527-86  
29 34.4 0.9 738 4 US-09-004-838-62  
30 34.2 0.9 1003 4 US-09-221-017B-395  
31 34.2 0.9 2694 1 US-08-465-995A-1  
32 34.2 0.9 2694 2 US-08-465-994C-1  
33 34.2 0.9 2694 2 US-08-966-145-1  
34 34.2 0.9 2760 1 US-08-101-593-1  
35 33.8 0.9 258 4 US-09-345-882-21  
36 33.8 0.9 1867 4 US-09-377-557-3  
37 33.8 0.9 2889 4 US-09-134-001C-1660  
38 33.6 0.9 2870 4 US-09-134-078-59  
39 33.6 0.9 3707 1 US-08-118-101A-1  
40 33.6 0.9 10223 4 US-08-961-527-73  
41 33.4 0.9 1939 1 US-08-232-188-3  
42 33.4 0.9 1939 3 US-08-970-725-3  
43 33.4 0.9 1939 5 PCT-US95-02708-3  
44 33.4 0.9 2026 3 US-08-755-587-26  
45 33.2 0.9 3636 4 US-09-134-001C-1983

Query Match				Score 3731; DB 4; Length 3731;			
Best Local Similarity				100.0%; Pred. No. 0;			
Matches 3731; Conservative				0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	GAATAATCTCTTACTTACTTACCAGGATGACTCATCACCCTCCCTCAAGTCTTTGTGT	60				
Db	1	GAATAATCTCTTACTTACTTACCAGGATGACTCATCACCCTCCCTCAAGTCTTTGTGT	60				
Qy	61	GTGTGATATAAATTTGGTGTGCTTCAGTTTCAGTCTACTACTGCTGGTAGTTTAT	120				
Db	61	GTGTGATATAAATTTGGTGTGCTTCAGTTTCAGTCTACTACTGCTGGTAGTTTAT	120				
Qy	121	TTTGCAATACTTCAGGGGATTCACAGTGTGTAGCATTTGAAAGTGGAACTGCACAT	180				
Db	121	TTTGCAATACTTCAGGGGATTCACAGTGTGTAGCATTTGAAAGTGGAACTGCACAT	180				
Qy	181	GGATTTGGCTACATGGGAAGCAATTCAGTTTTCGGATTCCTTCTCTCTG	240				
Db	181	GGATTTGGCTACATGGGAAGCAATTCAGTTTTCGGATTCCTTCTCTCTG	240				
Qy	241	GGAGTGGTTAAGTCATCTTAGAGAAATACAGGATTTGGAACGTGATGTCATTTGA	300				
Db	241	GGAGTGGTTAAGTCATCTTAGAGAAATACAGGATTTGGAACGTGATGTCATTTGA	300				
Qy	301	GGTTAAACATCCAAAGGAGATCTAGAGCATTTGCCAAGTTCAATTTGCCGACAACT	360				
Db	301	GGTTAAACATCCAAAGGAGATCTAGAGCATTTGCCAAGTTCAATTTGCCGACAACT	360				
Qy	361	AAGTGTGCAAAATCATCACTTTGGCTAATAACAGCTGTATTTGGCTCTCTTATTT	420				
Db	361	AAGTGTGCAAAATCATCACTTTGGCTAATAACAGCTGTATTTGGCTCTCTTATTT	420				
Qy	421	GAAGGCTTGGAAATGAAACTGATATTGCCAACTGCCGGCATATGTGGATCAGATGA	480				
Db	421	GAAGGCTTGGAAATGAAACTGATATTGCCAACTGCCGGCATATGTGGATCAGATGA	480				
Qy	481	TGGATTAACCTTGAATTCGGATGTCAGATATCAGATGACAACTTTCGAGTGTGGAG	540				
Db	481	TGGATTAACCTTGAATTCGGATGTCAGATATCAGATGACAACTTTCGAGTGTGGAG	540				
Qy	541	TACAGAAGTTCAATTCATTTGSCATTTGGATTGAAGAAATTTTCTTTTATCTAG	600				
Db	541	TACAGAAGTTCAATTCATTTGSCATTTGGATTGAAGAAATTTTCTTTTATCTAG	600				
Qy	601	TGGTTACGCTGACTATAAATTCAGCTTTCATATGAATAATATATGACAGTTCTGCTCA	660				
Db	601	TGGTTACGCTGACTATAAATTCAGCTTTCATATGAATAATATATGACAGTTCTGCTCA	660				
Qy	661	TCGTCCATATGGTCAAAATGCTCAGTTTCTCCTCATACAGTTATTTGGTCTCTCGGAT	720				
Db	661	TCGTCCATATGGTCAAAATGCTCAGTTTCTCCTCATACAGTTATTTGGTCTCTCGGAT	720				
Qy	721	CTATAAGAGACTTGAAACCTCCTGTATAGCTTCTTTAAGGAAACCTCTGATCAGTG	780				
Db	721	CTATAAGAGACTTGAAACCTCCTGTATAGCTTCTTTAAGGAAACCTCTGATCAGTG	780				
Qy	781	GGTGAGACACAGATTTCCCTCATCTCTGATAGGCTATCTTCTAGCTTATGTTTGA	840				
Db	781	GGTGAGACACAGATTTCCCTCATCTCTGATAGGCTATCTTCTAGCTTATGTTTGA	840				
Qy	841	GTTCCTGAGGGGTGCTGCTTCCAAATTTTCGAGGAAAGTTTTTCCACTATCCGAAAG	900				
Db	841	GTTCCTGAGGGGTGCTGCTTCCAAATTTTCGAGGAAAGTTTTTCCACTATCCGAAAG	900				
Qy	901	TGAAAAAATATTTACTTTACAGACTGGTTTCACCTTTTTCGCTCTCAAAAATCGGCTCT	960				
Db	901	TGAAAAAATATTTACTTTACAGACTGGTTTCACCTTTTTCGCTCTCAAAAATCGGCTCT	960				
Qy	961	GGTTCCCAATGCTCCAGCTCCGGAAGAAATTCAAATTCCTACAGATTTTGTTCAAAAT	1020				
Db	961	GGTTCCCAATGCTCCAGCTCCGGAAGAAATTCAAATTCCTACAGATTTTGTTCAAAAT	1020				
Qy	1021	TAGTCTTCTTGGTACAGCATGGATGATACCTGGGCCAGCATTTAAATGCTACTTTTCCG	1080				

Db	1021	TAGTCTTCTTGGTACAGCATGGATGATACCTGGGCCACCATTAATGCTACTTTTCCG	1080				
Qy	1081	ATTAGTTGATCCTCGAAGAGAAATGTGGCATGTGATGAGCATGCTTAGAAGAACTGTA	1140				
Db	1081	ATTAGTTGATCCTCGAAGAGAAATGTGGCATGTGATGAGCATGCTTAGAAGAACTGTA	1140				
Qy	1141	CTATATAAGAGAGTCTGTATGATCCCGTGGGTGGCTCACTGAGCAGTATGATGGTA	1200				
Db	1141	CTATATAAGAGAGTCTGTATGATCCCGTGGGTGGCTCACTGAGCAGTATGATGGTA	1200				
Qy	1201	TCTCAAGGTTAGACAACCTCCAAAATCTCCGTCCTCACTTTAGATGATGGTGGTGA	1260				
Db	1201	TCTCAAGGTTAGACAACCTCCAAAATCTCCGTCCTCACTTTAGATGATGGTGGTGA	1260				
Qy	1261	TCTTAAAGGTCCTAGTAAACACCATGCAAAAGTTTATTTTGTGTGTCAGAGTTAAATGT	1320				
Db	1261	TCTTAAAGGTCCTAGTAAACACCATGCAAAAGTTTATTTTGTGTGTCAGAGTTAAATGT	1320				
Qy	1321	TTTCCAATCGGTTCTCCGCAATTTATCTGAAGACATAGATAACTTTCTCGTGTCTTT	1380				
Db	1321	TTTCCAATCGGTTCTCCGCAATTTATCTGAAGACATAGATAACTTTCTCGTGTCTTT	1380				
Qy	1381	TGTTTATGAGGAGTGGGAAACTGTATCTACAGACTTATACCAAGCAAGTACTGG	1440				
Db	1381	TGTTTATGAGGAGTGGGAAACTGTATCTACAGACTTATACCAAGCAAGTACTGG	1440				
Qy	1441	AACTGGTGTGACGACAAACATCTATGAGAGTCTTATCAACTCTCGCGAAAGGCTTTGT	1500				
Db	1441	AACTGGTGTGACGACAAACATCTATGAGAGTCTTATCAACTCTCGCGAAAGGCTTTGT	1500				
Qy	1501	AATTTGGTGAATAAAATTTGAATTTCTTTCGATTTTCATCGAGCCAGTTGCGGGATAATTC	1560				
Db	1501	AATTTGGTGAATAAAATTTGAATTTCTTTCGATTTTCATCGAGCCAGTTGCGGGATAATTC	1560				
Qy	1561	AGTGTGGATGTTTGCATCAAGACCTGGCTTACTGCAAAATGATATAAGAGCTTGGATGG	1620				
Db	1561	AGTGTGGATGTTTGCATCAAGACCTGGCTTACTGCAAAATGATATAAGAGCTTGGATGG	1620				
Qy	1621	TGATTTTTCGAGATCAAGATGTCGAAAATATGCTGCCAGACTGCTCAATCTTTTGG	1680				
Db	1621	TGATTTTTCGAGATCAAGATGTCGAAAATATGCTGCCAGACTGCTCAATCTTTTGG	1680				
Qy	1681	TTCTCCACAGAGACTTTGAGTGTCTTAGGCATGAGATGAAGTTATTCGCGATGATAA	1740				
Db	1681	TTCTCCACAGAGACTTTGAGTGTCTTAGGCATGAGATGAAGTTATTCGCGATGATAA	1740				
Qy	1741	GGTTCATGGAACCCAGCTATGCTTTTCTGATGGAAATGGTAAATATCTGGTACTTTGC	1800				
Db	1741	GGTTCATGGAACCCAGCTATGCTTTTCTGATGGAAATGGTAAATATCTGGTACTTTGC	1800				
Qy	1801	TCATAGAGTTGCTCAAAATGTGGCTTCAATATACCCCATCTGCTTCCAGATTCGTTA	1860				
Db	1801	TCATAGAGTTGCTCAAAATGTGGCTTCAATATACCCCATCTGCTTCCAGATTCGTTA	1860				
Qy	1861	TGTTGGATATAAAGTGTGTTGGTGTGATCCGGATTCATCAATGAAGTTGCTTTGAG	1920				
Db	1861	TGTTGGATATAAAGTGTGTTGGTGTGATCCGGATTCATCAATGAAGTTGCTTTGAG	1920				
Qy	1921	AAAGCAGTGTGAAATATGAATGATCAGACAACTAAGTAGATGCTTGGATGGAGCAA	1980				
Db	1921	AAAGCAGTGTGAAATATGAATGATCAGACAACTAAGTAGATGCTTGGATGGAGCAA	1980				
Qy	1981	ATATCAGCTTCTTATCTTAATTCGTCAACTGATTAAGTCTTGTCTACACTTGGAGTAA	2040				
Db	1981	ATATCAGCTTCTTATCTTAATTCGTCAACTGATTAAGTCTTGTCTACACTTGGAGTAA	2040				
Qy	2041	AGATGAAGTTCTCGAACAGCAAAAGGAGCTGTAGATCAGCTTGATGCTATCTTCA	2100				
Db	2041	AGATGAAGTTCTCGAACAGCAAAAGGAGCTGTAGATCAGCTTGATGCTATCTTCA	2100				
Qy	2101	TGATTTCTTTGAAGGACAGAGGCTTTGGAATTTGATGCTCTCTGGAGAGAACATAAT	2160				

Db 2101 TGATCTTTTGAAGCAGCAGGAGGCTTTGGAAATGATGCTCTCTGGAGAGAACTATAT 2160  
Qy 2161 TCTCAAGCAATGCTAAACTGTGTTTATAAGCCCTGATGCTGAGCCCTTTCTTTCAATGAT 2220  
Db 2161 TCTCAAGCAATGCTAAACTGTGTTTATAAGCCCTGATGCTGAGCCCTTTCTTTCAATGAT 2220  
Qy 2221 GTTCGAACCTTCGCGCATCCCAAGTTCCTGATTTCCGAGCTAGATCAAGATATTTAT 2280  
Db 2221 GTTCGAACCTTCGCGCATCCCAAGTTCCTGATTTCCGAGCTAGATCAAGATATTTAT 2280  
Qy 2281 TCCAAATGGAAACAAATGATGGATGTTTGGATGAATCCAGAACCTTGGAAATATGCTCA 2340  
Db 2281 TCCAAATGGAAACAAATGATGGATGTTTGGATGAATCCAGAACCTTGGAAATATGCTCA 2340  
Qy 2341 GGTGTTTGTTCAGTTTACTGCTGACATGGAGATTTCTGACCATTTACATCCATT 2400  
Db 2341 GGTGTTTGTTCAGTTTACTGCTGACATGGAGATTTCTGACCATTTACATCCATT 2400  
Qy 2401 TAATAACAGCAGATCCCAACACGATTAATTCATCTCAAGGGAATGTGTTTGTGCAAA 2460  
Db 2401 TAATAACAGCAGATCCCAACACGATTAATTCATCTCAAGGGAATGTGTTTGTGCAAA 2460  
Qy 2461 AAATCCATGCTTGATGCTGCTGATATTCGTTTAAAGGCTGTAATGTTGAGGCT 2520  
Db 2461 AAATCCATGCTTGATGCTGCTGATATTCGTTTAAAGGCTGTAATGTTGAGGCT 2520  
Qy 2521 GCACACATGCTAGATGTTGTTGATTCCTCCAGAACGAAAGACCTCATCCGAATGA 2580  
Db 2521 GCACACATGCTAGATGTTGTTGATTCCTCCAGAACGAAAGACCTCATCCGAATGA 2580  
Qy 2581 ATGTTCTGGAGATGATTTGGATGGGATATCTACTTTTGTGTTGGGATCAAGACATGAT 2640  
Db 2581 ATGTTCTGGAGATGATTTGGATGGGATATCTACTTTTGTGTTGGGATCAAGACATGAT 2640  
Qy 2641 CCGGCAAGGCAAGTCCAGCGGATGGAAATCCCTCCAGCACCCAGCATACAGTTGGACCA 2700  
Db 2641 CCGGCAAGGCAAGTCCAGCGGATGGAAATCCCTCCAGCACCCAGCATACAGTTGGACCA 2700  
Qy 2701 TGATGTCACAAATGAGGAAGTTGAAGTACTTCCACCACTATATTTGTAATGACAGTTT 2760  
Db 2701 TGATGTCACAAATGAGGAAGTTGAAGTACTTCCACCACTATATTTGTAATGACAGTTT 2760  
Qy 2761 GGGATATACCAATGCCATGCTGATGTTTGCAGACAGAACCTCATATGCCATGAG 2820  
Db 2761 GGGATATACCAATGCCATGCTGATGTTTGCAGACAGAACCTCATATGCCATGAG 2820  
Qy 2821 TGATCCATGCAAAAACCTGCTGAGCTCTTTTCAATTCGAGTGGACCTTCCAAAGACTGG 2880  
Db 2821 TGATCCATGCAAAAACCTGCTGAGCTCTTTTCAATTCGAGTGGACCTTCCAAAGACTGG 2880  
Qy 2881 TGTTCCTGCTGAATACCATCTAGTTGCGCCCTTAAAGATACCCAGACTTCATGATAA 2940  
Db 2881 TGTTCCTGCTGAATACCATCTAGTTGCGCCCTTAAAGATACCCAGACTTCATGATAA 2940  
Qy 2941 CCGGCAAGCAGCAGCTATATCTCAGAAAGAGTTTATTTGAAAGCTTTTCCAGGAAAGTAA 3000  
Db 2941 CCGGCAAGCAGCAGCTATATCTCAGAAAGAGTTTATTTGAAAGCTTTTCCAGGAAAGTAA 3000  
Qy 3001 GGACAAAGCAGCTCAGGCTAGCTTATGCGCACCTTCCAGAGAGATGTTGCAAGGATC 3060  
Db 3001 GGACAAAGCAGCTCAGGCTAGCTTATGCGCACCTTCCAGAGAGATGTTGCAAGGATC 3060  
Qy 3061 ATATGATGCTGATATGGAAGTTGATGTTTGAAGATTACATTACCAAGAGCTTTTGACTA 3120  
Db 3061 ATATGATGCTGATATGGAAGTTGATGTTTGAAGATTACATTACCAAGAGCTTTTGACTA 3120  
Qy 3121 CAAAATGAATATGACAAACAGCTGGTAAATTTAATGACTACTATGGCATATAAAACAGA 3180  
Db 3121 CAAAATGAATATGACAAACAGCTGGTAAATTTAATGACTACTATGGCATATAAAACAGA 3180  
Qy 3181 GCCTGAATACTTAGTGTGCTGATATGAAGGATCAAAAACCTTTGACCCGACAAAAGA 3240  
Db 3181 GCCTGAATACTTAGTGTGCTGATATGAAGGATCAAAAACCTTTGACCCGACAAAAGA 3240

Qy 3241 TGCTGAGGCCATTAGTGTGCTGAGGGCCCTTGAGGAAGGAGGCAAGAGCCCTGGTTCAA 3300  
Db 3241 TGCTGAGGCCATTAGTGTGCTGAGGGCCCTTGAGGAAGGAGGCAAGAGCCCTGGTTCAA 3300  
Qy 3301 GAGGCGTAATGATATAGATGACATGTTACCAAGGCTTCGGCTTGGTACACGTTACATA 3360  
Db 3301 GAGGCGTAATGATATAGATGACATGTTACCAAGGCTTCGGCTTGGTACACGTTACATA 3360  
Qy 3361 TCATCTACATATTTGGGTTGCTACAAATCAGGGGTTGAAAAGAGCTCATTTCATTAGCTT 3420  
Db 3361 TCATCTACATATTTGGGTTGCTACAAATCAGGGGTTGAAAAGAGCTCATTTCATTAGCTT 3420  
Qy 3421 TCCTCGTGTGTTTANGACCACTATCCAGATTAAGAGGACCAAGACGCTAACAGGCC 3480  
Db 3421 TCCTCGTGTGTTTANGACCACTATCCAGATTAAGAGGACCAAGACGCTAACAGGCC 3480  
Qy 3481 AGTTCTCAACTTGTCTACTCTCTCAGGGCTCAACTGAGTACAGATTTAGTGAATGAGA 3540  
Db 3481 AGTTCTCAACTTGTCTACTCTCTCAGGGCTCAACTGAGTACAGATTTAGTGAATGAGA 3540  
Qy 3541 TTCCAGTCGAGCGTTAAGCTGATATATATATATATATATATATATATATATATATATAT 3600  
Db 3541 TTCCAGTCGAGCGTTAAGCTGATATATATATATATATATATATATATATATATATATAT 3600  
Qy 3601 TGTATGCTATGTTGACTACCTTTTCTTTAAACTGATGAAGCTGCAACATATATGC 3660  
Db 3601 TGTATGCTATGTTGACTACCTTTTCTTTAAACTGATGAAGCTGCAACATATATGC 3660  
Qy 3661 AGTACTCTAAGAACAGATGTACAGCTAAGTACTATATATATATATATATATATATATAT 3720  
Db 3661 AGTACTCTAAGAACAGATGTACAGCTAAGTACTATATATATATATATATATATATATAT 3720  
Qy 3721 TTTCTTCTTAAA 3731  
Db 3721 TTTCTTCTTAAA 3731

RESULT 2  
US-08-100-874-1/c  
; Sequence 1, Application US/08100874  
; Patent No. 5498533  
; GENERAL INFORMATION:  
; APPLICANT: Poovaiyah, B. W.  
; APPLICANT: Takezawa, D.  
; APPLICANT: Han, T. J.  
; APPLICANT: An, G. H.  
; TITLE OF INVENTION: Control of Growth and Development of  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harness, Dickey & Pierce  
; STREET: P.O. Box 828  
; CITY: Bloomfield Hills  
; STATE: MI  
; COUNTRY: USA  
; ZIP: 48303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/100,874  
; FILING DATE: July 30, 1993  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, Deann F.  
; REFERENCE/DOCKET NUMBER: 7555-00004  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (313) 641-1600  
; TELEFAX: (313) 641-0270  
; TELEX: 267637 Harness UR



```
, LOCATION: 93714
, OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT
, FEATURE:
, NAME/KEY: allele
, LOCATION: 97122
, OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 97152
, OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 99098
, OTHER INFORMATION: 5-130-257 : polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 99117
, OTHER INFORMATION: 5-130-276 : polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 103806
, OTHER INFORMATION: 5-131-395 : polymorphic base A or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 106940
, OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
, FEATURE:
, NAME/KEY: allele
, LOCATION: 108106
, OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A
, FEATURE:
, NAME/KEY: allele
, LOCATION: 108149
, OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTTT
, FEATURE:
, NAME/KEY: allele
, LOCATION: 108308
, OTHER INFORMATION: 5-135-357 : polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 108471
, OTHER INFORMATION: 5-136-174 : polymorphic base C or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 134134
, OTHER INFORMATION: 5-140-120 : polymorphic base C or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 134362
, OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
, FEATURE:
, NAME/KEY: allele
, LOCATION: 134374
, OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
, FEATURE:
, NAME/KEY: allele
, LOCATION: 146328
, OTHER INFORMATION: 5-143-84 : polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 146345
, OTHER INFORMATION: 5-143-101 : polymorphic base A or C
, FEATURE:
, NAME/KEY: allele
, LOCATION: 150329
, OTHER INFORMATION: 5-145-24 : polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 160031
, OTHER INFORMATION: 5-148-352 : polymorphic base G or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 72771..72817
, OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
, FEATURE:
, NAME/KEY: allele
, LOCATION: 72771..72817
, OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
, FEATURE:
, NAME/KEY: allele
, LOCATION: 88050..88096
, OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
, FEATURE:
, NAME/KEY: allele
, LOCATION: 88050..88096
, OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
, FEATURE:
, NAME/KEY: allele
, LOCATION: 90819..90865
, OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
, FEATURE:
, NAME/KEY: allele
, LOCATION: 90819..90865
, OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
, FEATURE:
, NAME/KEY: allele
, LOCATION: 93690..93736
, OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
, FEATURE:
, NAME/KEY: allele
, LOCATION: 93690..93736
, OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
, FEATURE:
, NAME/KEY: allele
, LOCATION: 97099..97145
, OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
, FEATURE:
, NAME/KEY: allele
, LOCATION: 97099..97145
, OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
, FEATURE:
, NAME/KEY: allele
, LOCATION: 97130..97177
, OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
, FEATURE:
, NAME/KEY: allele
, LOCATION: 97130..97177
, OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
, FEATURE:
, NAME/KEY: allele
, LOCATION: 99075..99121
, OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
, FEATURE:
, NAME/KEY: allele
, LOCATION: 99075..99121
, OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
, FEATURE:
, NAME/KEY: allele
, LOCATION: 99094..99140
, OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
, FEATURE:
, NAME/KEY: allele
, LOCATION: 99094..99140
, OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
, FEATURE:
, NAME/KEY: allele
, LOCATION: 103783..103828
, OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
, FEATURE:
, NAME/KEY: allele
, LOCATION: 103783..103828
, OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
, FEATURE:
, NAME/KEY: allele
, LOCATION: 106918..106966
, OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
```

```

, FEATURE:
, NAME/KEY: allele
, LOCATION: 105918..106966
, OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
, FEATURE:
, NAME/KEY: allele
, LOCATION: 108084..108130
, OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
, FEATURE:
, NAME/KEY: allele
, LOCATION: 108084..108130
, OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
, FEATURE:
, NAME/KEY: allele
, LOCATION: 108127..108177
, OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
, FEATURE:
, NAME/KEY: allele
, LOCATION: 108127..108177
, OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
, FEATURE:

```

Query Match	1.18;	Score 40;	DB 4;	Length 162450;
Best Local Similarity	47.6%;	Pred. No. 2.1;		
Matches 118;	Conservative 0;	Mismatches 130;	Indels 0;	Gaps 0;
QY 3032	ACCTTCACAAGAGATGTTGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGCAATTT	3091		
Db 114325	AACCTTTCAGAGATGAACGAATAAAGAAGAAGATGAAGATGATGAAGAAGCAGAAGAG	114384		
QY 3092	GRAGATTACATTCGACGAGCCTTTTGACTACAAACTCAATATCACACACAGCTGGGTAT	3151		
Db 114385	GAGGAGGAGGGAAGAAGAAGAGGATGAAGATGATGATGACACAAATGAGGAAGAG	114444		
QY 3152	TTAATGGACTACTATGGCATAAAACACAGAGGCTGAAATACTTATGGTGGCATATTGAAG	3211		
Db 114445	GAGTTTGGTGTCTCCACAGGCATGAAAGTCCCAAGTGCGGTATGACAGAGGGAAAAAT	114504		
QY 3212	GCATCAAAAACCTTTTCACGCGAGAAAAGATCCTGAGGCCATTAGTTGCTGTGAGGGCC	3271		
Db 114505	CAAAAATGTATGAAGCTAGTATATTAAGATTCTGATCTCGAAGGTGGAGNGTCTCTTAC	114564		
QY 3272	TTGAGAA	3279		
Db 114565	TTGGTGCA	114572		

```

RESULT 5
US-08-676-967-2
; Sequence 2, Application US/08676967
; Patent No. 5747317
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,967
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A

```

```

; REGISTRATION NUMBER: 36 627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
;
US-08-676-967-2

Query Match 1.0%; Score 36.8; DB 1; Length 2277;
Best Local Similarity 27.4%; Pred. No. 1.8;
Matches 113; Conservative 66; Mismatches 232; Indels 2; Gaps

QY 2983 TTCCGCTGAATACCATCTCAGTTGCGCCCTAAAGATAACCACAGATTCATGGTAAAGC 2942
DB 154 TTTWSNATGTTNGARGAYGTGCAEMGNGGNTTNAARGARATHACNCTTTGARGGNTGY 213
QY 2943 CGGACAAGACCGACTATCTCAGAAAGAGTTATTGGAAAGCTTTTCAGGAAATGAGG 3002
DB 214 AARATHAAYGTNACNGTNGCNAARAARAARYTNMGNAAYAAACNAARGARAAAGGNAAR 273
QY 3003 ACAAAGCACCTCAGGCTAGCTCTATCGCGACCTTCACAAGAGATGTTGCAAGGAGATCAT 3062
DB 274 AAYGARAAYWSNGARTGYCCNARAARGARCCNARGCNARAARGCNAAARGTNGCNGAY 333
QY 3063 ATGATGCTGATATGGAAGTTGATGGATTGGAAGATACATTGACGAAGCTTTTCACTACA 3122
DB 334 AARAARGCNGMNTTNATHMGNRAAYTNWSNTTYAARTGYWSNGAR--GAYGAYTTNA 391
QY 3123 AAATCGTAATATGACAACAGCTGGGTAAATTTAATGGACTACTATGGCATATAAAACAGAGG 3182
DB 392 ARACNGTNTTYGCNCARTTYGGNGCNGTNTNGARGTNAAYATHCNMGNAARCCNGAYG 451
QY 3183 CTGAAATACHTAGTGGTGGCATATGCAAGGCATCAAAACATTTTGACCGCAGAAAGATG 3242
DB 452 GNAARATGTMGNGGNTTYGGNTTYGTNCARTTYAARAAYTNTTNGARCGNCGNARGCNY 511
QY 3243 CTGAGGCCATTAGTGTGTTGTTGTCGAGGGCCCTTGAGGAAGGAGGCAAGACCTGG 3295
DB 512 TNAARGGNATCAAYATGAARGARATHAARGMNGMGNACNGTNGCTNGCTNGAYTGG 564

```

RESULT 6  
US-08-676-974-2  
; Sequence 2, Application US/08676974  
; Patent No. 5770422  
; GENERAL INFORMATION:  
; APPLICANT: COLLINS, KATHLEEN  
; TITLE OF INVENTION: Human Telomerase  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Science & Technology Law Group  
; STREET: 268 Bush Street, Suite 3200  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/676,974  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman Ph.D., Richard A



REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: UCB96-055  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415)343-4341  
TELEFAX: (415)343-4342  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2277 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
US-09-098-487-2

Query Match 1.0%; Score 36.8; DB 2; Length 2277;  
Best Local Similarity 27.4%; Pred. No. 1.8;  
Matches 113; Conservative 66; Mismatches 232; Indels 2; Gaps 1;

QY 2883 TTCCCGCTGAATACCATCTCAGTTGGCGCTAAAGAAATACCCAGACTTCATGGATAAGC 2942  
DB 154 TTWSNATGTYNGARGAYTNCARMGNCNTYNAARGARATHACNACNTTYGARGGNTGY 213  
QY 2943 CGGACAGACCCAGCTATATCTCAGAAAGAGTTATTGGAAGCTTTTCAGGAAAGTGAAG 3002  
DB 214 AARATHAAYGTNACNGTNGCNAARARARYTNMGNAAYAAACNAARGAARGGNAAR 273  
QY 3003 ACAAGACACCTCAGGCTAGCTCTATCGGACCTTCAAGAGATGTTGCAAGGAGATCAT 3062  
DB 274 AAYGARAAYWSNGARTGYCCNARARARGCCNARARCCNARARCCNARARCCNGAY 333  
QY 3063 ATGATGCTGATATGGAAGTTGATGATTGAAGATTACATTCACCAAGCTTTTGACTACA 3122  
DB 334 AARAAGCNGMNTYNAATHMGNAAYTNWNTTYAARTGYWSNGAR--GAYGAYTTNA 391  
QY 3123 AAACCTGAATATGACACCAAGCTGGTGAATTAATGAGCTACTATGGCATAAAAACAGAG 3182  
DB 392 ARACNGTNTTYGNCARTTYGGNGCNGTNTYNGARGTNAAYATHCCNMGNARCCNGAYG 451  
QY 3183 CTGAAATACCTTAGTGGTGCATATGAAGCATCAAAACTTTTGACCGCAGAAAAAGATG 3242  
DB 452 GNAARATGGMGNGTNTTYGNTTNCARTTYAARAAYTYNTYNGARGCNGGNAARGCNY 511  
QY 3243 CTGAGGCCATATAGTGTGCTGTGAGGCGCTTGAGGAGGAGGAGGAGGAGGAGGAGG 3295  
DB 512 TNAARGGNATGAAYATGAARGARATHAARGGNMGNCNGTNGCTNGAYTGG 564

RESULT 8  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435

REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: UCB96-055  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415)343-4341  
TELEFAX: (415)343-4342  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2277 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
US-08-676-974-2

Query Match 1.0%; Score 36.8; DB 1; Length 2277;  
Best Local Similarity 27.4%; Pred. No. 1.8;  
Matches 113; Conservative 66; Mismatches 232; Indels 2; Gaps 1;

QY 2883 TTCCCGCTGAATACCATCTCAGTTGGCGCTAAAGAAATACCCAGACTTCATGGATAAGC 2942  
DB 154 TTWSNATGTYNGARGAYTNCARMGNCNTYNAARGARATHACNACNTTYGARGGNTGY 213  
QY 2943 CGGACAGACCCAGCTATATCTCAGAAAGAGTTATTGGAAGCTTTTCAGGAAAGTGAAG 3002  
DB 214 AARATHAAYGTNACNGTNGCNAARARARYTNMGNAAYAAACNAARGAARGGNAAR 273  
QY 3003 ACAAGACACCTCAGGCTAGCTCTATCGGACCTTCAAGAGATGTTGCAAGGAGATCAT 3062  
DB 274 AAYGARAAYWSNGARTGYCCNARARARGCCNARARCCNARARCCNARARCCNGAY 333  
QY 3063 ATGATGCTGATATGGAAGTTGATGATTGAAGATTACATTCACCAAGCTTTTGACTACA 3122  
DB 334 AARAAGCNGMNTYNAATHMGNAAYTNWNTTYAARTGYWSNGAR--GAYGAYTTNA 391  
QY 3123 AAACCTGAATATGACACCAAGCTGGTGAATTAATGAGCTACTATGGCATAAAAACAGAG 3182  
DB 392 ARACNGTNTTYGNCARTTYGGNGCNGTNTYNGARGTNAAYATHCCNMGNARCCNGAYG 451  
QY 3183 CTGAAATACCTTAGTGGTGCATATGAAGCATCAAAACTTTTGACCGCAGAAAAAGATG 3242  
DB 452 GNAARATGGMGNGTNTTYGNTTNCARTTYAARAAYTYNTYNGARGCNGGNAARGCNY 511  
QY 3243 CTGAGGCCATATAGTGTGCTGTGAGGCGCTTGAGGAGGAGGAGGAGGAGGAGGAGG 3295  
DB 512 TNAARGGNATGAAYATGAARGARATHAARGGNMGNCNGTNGCTNGAYTGG 564

RESULT 7  
US-09-098-487-2  
Sequence 2, Application US/09098487  
Patent No. 5917025  
GENERAL INFORMATION:  
APPLICANT: COLLINS, Kathleen  
TITLE OF INVENTION: Human Telomerase  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Science & Technology Law Group  
STREET: 268 Bush Street, Suite 3200  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/098,487  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman Ph.D., Richard A

APPLICANT: Rosenberg, Martin  
APPLICANT: Ward, Judith  
TITLE OF INVENTION: No. 5348582el Prokaryotic Polynucleotides,  
TITLE OF INVENTION: Polypeptides and Their Uses  
NUMBER OF SEQUENCES: 534  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/936.165A  
FILING DATE: 24-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/027,032  
FILING DATE: 24-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimml, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 194:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3191 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-08-936-165A-194

Query Match 1.0%; Score 36.4; DB 4; Length 3191;  
Best Local Similarity 49.0%; Pred. No. 2.8;  
Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 260 TTGAGAAATATACAGATATGGAACCTGATGTCATTTGGAGTTAAACAGTCCAAAGGA 319  
DB 2823 TTAAGAAACACATATTATGAAGATGTTGGACACCGGAGATATTGAAAAAATTTAT 2882  
QY 320 GGATCTAGAGCATTTGCCAAAGTTCAATTTGCCGACACATAGTCTGACAAAAATCATC 379  
DB 2883 CGNCTAATCGTGGTCAATATATATGTTGTGCGAGATAAAAGAAACAAAGGATTT 2942  
QY 380 ACTTTGCTTAATACAGGCTGATTTTGGCTCTCTTATTTCAGAGCTTGGGAAATGAAA 439  
DB 2943 ACCTTTCTTAAAGAAAGTCAGTATTTTGAACACTTGTACTTGTAGTGGATCAGNAAAT 3002  
QY 440 ACTGATATTGTC 451  
DB 3003 CCTGGTGGTGGC 3014

RESULT 10  
US-09-415-946-1/c  
Sequence 1, Application US/09415946  
Patent No. 6376751  
GENERAL INFORMATION:  
APPLICANT: Sung, Z. Renee  
APPLICANT: Aubert, Dominique  
APPLICANT: Chen, Lingjing  
APPLICANT: The Regents of the University of California  
TITLE OF INVENTION: Nucleic Acids That Control Reproductive Development in  
PLANTS

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/97/935.313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-Fls  
US-08-232-463-14

Query Match 1.0%; Score 36.6; DB 1; Length 7218;  
Best Local Similarity 10.1%; Pred. No. 3.8; Mismatches 223; Indels 0; Gaps 0;  
Matches 48; Conservative 204

QY 581 TTTTCTTTCTTTTACTAGTGTGCTGACTGACTATAAATCTACGCTTTCATATGAAAT 640  
DB 1074 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1133  
QY 641 ATATGACAGTGTGCTCCATGTCATATGTCATAATGCTCAATGCTCTCTCTCATACAG 700  
DB 1134 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1193  
QY 701 TTATTGGTCTCTCTCGGATCTATAGAGACTGAAACTCTCTGTATAGCTCTCTTAAG 760  
DB 1194 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1253  
QY 761 GAACTCTCTGATGATGAGTGGTGAGGACAAACAGATTTCCCTCCATCTTGGATAGGGCTA 820  
DB 1254 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1313  
QY 821 TCTCTAGCTATGTTGCGAGTCCGATGAGGCTGCTGCTTCCAAATTCGAGAAAGT 880  
DB 1314 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1373  
QY 881 TTCTTCCATGATGAGAGCTGAAACATATTTACATTTACAGACTGGTTTACCTTTTTC 940  
DB 1374 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1433  
QY 941 GTCTCTCAAAATCGGCTCTGTTCCCATGTCACGCTCCGAGAGGAATTCATATGCC 1000  
DB 1434 YGTACCAAAATCTCTATCTCTTAACTACTTGTGATAGATAGTATACAGTATGCT 1493  
QY 1001 TACAAGATTTTGTTCAAAATAGTCTTCTTGGTACACATGATGTCATACCTGGGC 1055  
DB 1494 ACATGCCGTTTTTGAACATGATAGTGGCTGTAGAGCGATGCTACGCTAGTC 1548

RESULT 9  
US-08-936-165A-194  
Sequence 194, Application US/08936165A  
Patent No. 6348582  
GENERAL INFORMATION:  
APPLICANT: Black, Michael  
APPLICANT: Burnham, Martin  
APPLICANT: Hodgson, John  
APPLICANT: Knowles, David  
APPLICANT: Lonetto, Michael  
APPLICANT: Nicholas, Richard  
APPLICANT: Pratt, Julie  
APPLICANT: Reichard, Richard

FILE REFERENCE: 018941-000110US  
CURRENT APPLICATION NUMBER: US/09/415,946  
CURRENT FILING DATE: 1999-10-08  
PRIOR APPLICATION NUMBER: US 09/169,696  
PRIOR FILING DATE: 1998-10-09  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 17341  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
OTHER INFORMATION: genomic DNA (Ecotype Columbia) from CD82 clone  
NAME/KEY: Promoter  
LOCATION: (1)..(3201)  
NAME/KEY: CDS  
LOCATION: Join(4241..4335, 4448..4623, 4704..4823, 4903..4956,  
OTHER INFORMATION: EMBRYONIC FLOWER 1 (EMF1)  
NAME/KEY: exon  
LOCATION: (3202)..(3265)  
OTHER INFORMATION: exon 1  
NAME/KEY: intron  
LOCATION: (3266)..(4159)  
NAME/KEY: exon  
LOCATION: (4160)..(4335)  
OTHER INFORMATION: exon 2  
NAME/KEY: intron  
LOCATION: (4336)..(4447)  
NAME/KEY: exon  
LOCATION: (4448)..(4623)  
OTHER INFORMATION: exon 3  
NAME/KEY: intron  
LOCATION: (4624)..(4703)  
NAME/KEY: exon  
LOCATION: (4704)..(4823)  
OTHER INFORMATION: exon 4  
NAME/KEY: intron  
LOCATION: (4824)..(4902)  
NAME/KEY: exon  
LOCATION: (4903)..(4956)  
OTHER INFORMATION: exon 5  
NAME/KEY: intron  
LOCATION: (4957)..(5045)  
NAME/KEY: exon  
LOCATION: (5046)..(6307)  
OTHER INFORMATION: exon 6  
NAME/KEY: intron  
LOCATION: (6308)..(6447)  
NAME/KEY: exon  
LOCATION: (6448)..(8065)  
OTHER INFORMATION: exon 7  
NAME/KEY: intron  
LOCATION: (8066)..(8300)  
NAME/KEY: exon  
LOCATION: (8301)..(8648)  
OTHER INFORMATION: exon 8  
US-09-415-946-1

Query Match 0.9%; Score 35.4; DB 4; Length 17341;  
Best Local Similarity 49.2%; Pred. No. 13;  
Matches 93; Conservative 0; Mismatches 96; Indels 0; Gaps 0;  
QY 1115 ATGAGCATGCTTAGAAGAACTGTACTATATAAAGGAGTGCTGTATGATGCCGTGAGG 1174  
DB 10360 ATTTGACCTAACTTTGTAAATCTTGGAGTAAAGAACAACTAGATCTTAAGAAGATG 10301  
QY 1175 TGGCTCACTGACGATGATGGTGTATCTCAAGGCTAGAACCTCCAAATCTCCGTCC 1234  
DB 10300 TTCTCTAGTATGGTCTGATGAGATCTCTTAGGAACAGAGCTTCCAGTATGACGGA 10241  
QY 1235 ATCACTTTAGATGGGTGGTGTATGTAAAGAGGTCTCTAGTAACACATGCCAAGTT 1294  
DB 10240 TGTGCTTGGAGGAAATTTGTGTGTCTTCAAGAGAAAGATCTGAACCTTCCAAAGT 10181

QY 1295 TATTTTGT 1303  
DB 10180 TACTTTGT 10172

RESULT 11  
US-09-134-001C-342  
; Sequence 342, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: LYNN Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 342  
; LENGTH: 1248  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-342

Query Match 0.9%; Score 35; DB 4; Length 1248;  
Best Local Similarity 51.6%; Pred. No. 4.3;  
Matches 80; Conservative 0; Mismatches 75; Indels 0; Gaps 0;  
QY 3040 AAGAGATGTTGCAAGGAGATCATATGATGCTGATATGAAGTTGATGATTTGAAGATTA 3099  
DB 813 ATGCTTTTACAGGTAGCTTAGAAATGACCATGGAGATGCAATGAAGTGTAGATT 872  
QY 3100 CATTCACGAAGCTTTTGACTACAAACTGAATATGACACAAAGCTGGTGAATTATGGA 3159  
DB 873 GCTAAACAGCCTTAGAAGTCGATATGCATATCATCAACCATTTGTTTAACTGA 932  
QY 3160 CTACTATGTCATAAAACAGAGGCTGAAATACTTA 3194  
DB 933 TTATATCGTATGAAGAGACTATGATCAATTA 967

RESULT 12  
US-08-911-445-4  
; Sequence 4, Application US/08911445  
; Patent No. 5876713  
; GENERAL INFORMATION:  
; APPLICANT: TAKEDA CHEMICAL INDUSTRIES, LTD.  
; TITLE OF INVENTION: NOVEL GLUTAMINE:FRUCTOSE-6-PHOSPHATE  
; TITLE OF INVENTION: AMIDOTRANSFERASE, ITS PRODUCTION AND USE  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FETHERSTONHAUGH & Co.  
; STREET: P.O. BOX 2999, STATION D  
; CITY: OTTAWA  
; STATE: ONTARIO  
; COUNTRY: CANADA  
; ZIP: K1P 5Y6  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/911,445  
; FILING DATE: 12-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 213944-1996  
; FILING DATE: 13-AUG-1996



Search completed: November 5, 2002, 23:30:56  
Job time : 710 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: November 5, 2002, 22:34:33 ; Search time 9169 Seconds  
(without alignments)  
11842.357 Million cell updates/sec  
Title: US-09-782-874-1  
Perfect score: 3731  
Sequence: 1 GAAATATCTTACTTACTT.....AGTTTCATCTTCTTCTTCTTAA 3731  
Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0  
Searched: 2054640 seqs, 14551402878 residues  
Word size : 0  
Total number of hits satisfying chosen parameters: 4109280  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Listing first 50 summaries  
Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_on.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_or.\*  
21: em\_ov.\*  
22: em\_pat.\*  
23: em\_ph.\*  
24: em\_pl.\*  
25: em\_ro.\*  
26: em\_sts.\*  
27: em\_un.\*  
28: em\_vl.\*  
29: em\_hgt\_hum.\*  
30: em\_hgt\_inv.\*  
31: em\_hgt\_other.\*  
32: em\_hgt\_mus.\*  
33: em\_hgt\_pln.\*  
34: em\_hgt\_rod.\*  
35: em\_hgt\_mam.\*  
36: em\_hgt\_vrt.\*  
37: em\_sy.\*  
38: em\_hgt\_hum.\*  
39: em\_hgt\_mus.\*  
40: em\_hgt\_pln.\*  
41: em\_hgt\_rod.\*  
42: em\_hgt\_mam.\*  
43: em\_hgt\_vrt.\*  
44: em\_sy.\*  
45: em\_hgt\_hum.\*  
46: em\_hgt\_mus.\*  
47: em\_hgt\_pln.\*  
48: em\_hgt\_rod.\*  
49: em\_hgt\_mam.\*  
50: em\_hgt\_vrt.\*

SD

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3731	100.0	3731	6	ARI45904 Sequence
2	3680	98.6	3731	8	LERDRP
3	94	2.5	3505	8	NTAC11576
4	73	2.0	4579	8	PHY011979
5	35	0.9	1838	8	ATH011977
6	30	0.8	908	6	I18797 Sequence
7	30	0.8	906	8	POTCAM
8	25	0.7	25	6	ARI45905
9	25	0.7	25	6	ARI45907
10	24	0.6	25	6	ARI45906
11	23	0.6	112247	2	AC095524 Rattus no
12	23	0.6	132699	8	AC006917
13	22	0.6	204160	2	AC113586
14	21	0.6	54683	2	AL732358_3
15	21	0.6	62512	2	AC103999
16	21	0.6	64179	9	AL627224
17	21	0.6	86478	2	AC022577
18	21	0.6	101706	2	AP004079
19	21	0.6	122685	2	AC099102
20	21	0.6	129226	9	HS960017
21	21	0.6	131310	2	AC126008
22	21	0.6	135458	2	AC117166
23	21	0.6	137723	9	HS060622
24	21	0.6	138931	2	AP004023
25	21	0.6	142234	9	AC110620
26	21	0.6	150055	2	AC108260
27	21	0.6	153258	2	AC128389
28	21	0.6	155859	2	AC121213
29	21	0.6	157759	2	AC114071
30	21	0.6	158158	2	AC034304
31	21	0.6	158371	9	AC011155
32	21	0.6	160083	2	AC129367
33	21	0.6	162600	2	AP001802
34	21	0.6	165699	2	AP001369
35	21	0.6	168763	2	AP000880
36	21	0.6	171196	2	AC126420
37	21	0.6	171853	9	AP002840
38	21	0.6	175245	2	AC109420
39	21	0.6	177034	9	AC009430
40	21	0.6	178229	2	AC109179
41	21	0.6	179025	2	AP001934
42	21	0.6	179281	2	AC083994
43	21	0.6	181265	2	AC113269
44	21	0.6	183951	2	AC096641
45	21	0.6	185148	9	AC084367
46	21	0.6	186386	2	AC125777
47	21	0.6	189748	2	AC009466
48	21	0.6	193279	9	AC055817
49	21	0.6	202781	2	AC055817
50	21	0.6	202781	2	AC055817

ALIGNMENTS

RESULT 1  
ARI45904  
LOCUS  
DEFINITION Sequence 1 from patent US 6218142.  
ACCESSION ARI45904  
VERSION ARI45904.1 GI:15109093  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

ARI45904 3731 bp DNA 1linear PAT 08-AUG-2001

Pred. No. is the number of results predicted by chance to have a

REFERENCE 1 (bases 1 to 3731)  
AUTHORS Wassenegeger,M., Riedel,L., Schiebel,W. and Sanger,H.L.  
TITLE Nucleic acid molecules encoding polypeptides having the enzymatic activity of an RNA-directed RNA polymerase (RDRP)  
JOURNAL Patent: US 6218142-A 1 17-APR-2001;  
FEATURES Location/Qualifiers  
source 1..3731  
BASE COUNT 1064 a 569 c 849 g 1149 t  
ORIGIN  
Query Match 100.0%; Score 3731; DB 6; Length 3731;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAATATTTCTTACTTACTTCCACAGGATGACTCATCTACCTCCCTCAAGTCCTTTGTGT 60  
Db 1 GAATATTTCTTACTTACTTCCACAGGATGACTCATCTACCTCCCTCAAGTCCTTTGTGT 60  
QY 61 GTTGATATAATAAATTTGGTTGTGCTTCAGTTTCAGTCACACTACCTGCTGGGTAGTTTTAT 120  
Db 61 GTTGATATAAATTTGGTTGTGCTTCAGTTTCAGTCACACTACCTGCTGGGTAGTTTTAT 120  
QY 121 TTTCGATAAATTCAGGGGATATCCAGTTGGTGTAGCATTTGAAAGTCGAATGCACCTT 180  
Db 121 TTTCGATAAATTCAGGGGATATCCAGTTGGTGTAGCATTTGAAAGTCGAATGCACCTT 180  
QY 181 GGAATTTGGCTACATGGGAAGACAATTCAGGTTTTCGGATTCCTTATCTCTCTGTC 240  
Db 181 GGAATTTGGCTACATGGGAAGACAATTCAGGTTTTCGGATTCCTTATCTCTCTGTC 240  
QY 241 GGAAGTGGTTAAGTCATCTTTAGAGAATATACAGGATATGGAACTGTAATGTCATTGGA 300  
Db 241 GGAAGTGGTTAAGTCATCTTTAGAGAATATACAGGATATGGAACTGTAATGTCATTGGA 300  
QY 301 GGTTAAACAGTCCAAAGGAGATCAGAGCATTTGCCAAAGTTCATTTGCCGACACAT 360  
Db 301 GGTTAAACAGTCCAAAGGAGATCAGAGCATTTGCCAAAGTTCATTTGCCGACACAT 360  
QY 361 AAGTGTGACAAAATCATCACTTTGGCTAATAACAGGCTGTATTTGGCTCTCTTATTT 420  
Db 361 AAGTGTGACAAAATCATCACTTTGGCTAATAACAGGCTGTATTTGGCTCTCTTATTT 420  
QY 421 GAAGCTTTGGGAATGAACCTGATATTTGCCAACTCGGGCATATGTCATCAGATGGA 480  
Db 421 GAAGCTTTGGGAATGAACCTGATATTTGCCAACTCGGGCATATGTCATCAGATGGA 480  
QY 481 TGGCATACCTTTGAATTCGGATGTCAGATATCAGATGACAAGTTTGCAGTGTGGGAAG 540  
Db 481 TGGCATACCTTTGAATTCGGATGTCAGATATCAGATGACAAGTTTGCAGTGTGGGAAG 540  
QY 541 TACAGAAGTTTCAATTCATTTGGCATTTGGATTGAAGAAATTTTTTCTTTTATCTAG 600  
Db 541 TACAGAAGTTTCAATTCATTTGGCATTTGGATTGAAGAAATTTTTTCTTTTATCTAG 600  
QY 601 TGGTTACGCTGACATATAACTTCAGCTTTCATATGAAATATATGCGAGCTTGTGCTCCA 660  
Db 601 TGGTTACGCTGACATATAACTTCAGCTTTCATATGAAATATATGCGAGCTTGTGCTCCA 660  
QY 661 TCGTCCATATGTTCAAAATGCTCAGTTTCCTCATACAGTTATTTGGTGTCTCTCGGAT 720  
Db 661 TCGTCCATATGTTCAAAATGCTCAGTTTCCTCATACAGTTATTTGGTGTCTCTCGGAT 720  
QY 721 CTATAAGAGACTTGAACCTCGTTATAGCTTCTTAAAGAAACCTCTGATCAGTG 780  
Db 721 CTATAAGAGACTTGAACCTCGTTATAGCTTCTTAAAGAAACCTCTGATCAGTG 780  
QY 781 GGTGAGGACACAGATTTCCCTCATCTTGGATAGGCTATCTTCTAGCTTATGTTTGA 840  
Db 781 GGTGAGGACACAGATTTCCCTCATCTTGGATAGGCTATCTTCTAGCTTATGTTTGA 840  
QY 841 GTTCCGTAGGGGTTGCTCTTCCAAATTTTCAGGAAAGTTTTTTCACATATCAGAAAG 900  
Db 841 GTTCCGTAGGGGTTGCTCTTCCAAATTTTCAGGAAAGTTTTTTCACATATCAGAAAG 900

Db 841 GTTCCGTAGGGGTTGCTCTTCCAAATTTTCAGGAAAGTTTTTTCACATATCAGAAAG 900  
QY 901 TGAACAAATATTTACTTTACAGACTGGTTTACCTTTTTCGCTCTCAAAAATCGGCTCT 960  
Db 901 TGAACAAATATTTACTTTACAGACTGGTTTACCTTTTTCGCTCTCAAAAATCGGCTCT 960  
QY 961 GGTTCCTCAATGTCCAGCTCCGGAAGAAATTTCAATTCCTCCTACAGATTTTGTTCAAAAT 1020  
Db 961 GGTTCCTCAATGTCCAGCTCCGGAAGAAATTTCAATTCCTCCTACAGATTTTGTTCAAAAT 1020  
QY 1021 TACTTCTTTTGGTACAGCATGGATGATACCTGGGCCAGCATTAATAGTCTACTTTTTCCG 1080  
Db 1021 TACTTCTTTTGGTACAGCATGGATGATACCTGGGCCAGCATTAATAGTCTACTTTTTCCG 1080  
QY 1081 ATTAGTGTATCTCGAAGGAGAAATGTGGCATGATGAGCATGCTTATAGAGAACTGTA 1140  
Db 1081 ATTAGTGTATCTCGAAGGAGAAATGTGGCATGATGAGCATGCTTATAGAGAACTGTA 1140  
QY 1141 CTATATAAGGAGTGTCTTATGATCCCGTGGTGGTCTACCTGAGCAGTATGATGGGTA 1200  
Db 1141 CTATATAAGGAGTGTCTTATGATCCCGTGGTGGTCTACCTGAGCAGTATGATGGGTA 1200  
QY 1201 TCTCAAGGTTAGACAACTCCAAAATCTCCGTCATCACTTTAGATGATGGGTGGTGA 1260  
Db 1201 TCTCAAGGTTAGACAACTCCAAAATCTCCGTCATCACTTTAGATGATGGGTGGTGA 1260  
QY 1261 TGTAAAGAGGTCCTAGTAAACCAATGCAAGTATTTTGTGGTCCAGAGGTTAAATGT 1320  
Db 1261 TGTAAAGAGGTCCTAGTAAACCAATGCAAGTATTTTGTGGTCCAGAGGTTAAATGT 1320  
QY 1321 TTCCAATCGGGTTCTCCGCAATTTATCTGAAGACATAGATAACTTTCTTCGTTCTCTTT 1380  
Db 1321 TTCCAATCGGGTTCTCCGCAATTTATCTGAAGACATAGATAACTTTCTTCGTTCTCTTT 1380  
QY 1381 TGTGTATGAGGAGTGGGAGAACTGTTTCTACAGACTTATTACCAAAAAGCAAGTACTGG 1440  
Db 1381 TGTGTATGAGGAGTGGGAGAACTGTTTCTACAGACTTATTACCAAAAAGCAAGTACTGG 1440  
QY 1441 AAGTGTGTGTCAGACAAACATCTATGAGAGGATCTTATCAACTCTCGGAAAAGGTTTGT 1500  
Db 1441 AAGTGTGTGTCAGACAAACATCTATGAGAGGATCTTATCAACTCTCGGAAAAGGTTTGT 1500  
QY 1501 AATTGTGATATAAAATTTTGAATTTCTTCATGAGGATCTTATCAACTCTCGGAAAAGTTC 1560  
Db 1501 AATTGTGATATAAAATTTTGAATTTCTTCATGAGGATCTTATCAACTCTCGGAAAAGTTC 1560  
QY 1561 AGTGTGATTTTGCATCAAGACCTGGCCCTTACTGCAATGATATAAGAGCTTGGATGG 1620  
Db 1561 AGTGTGATTTTGCATCAAGACCTGGCCCTTACTGCAATGATATAAGAGCTTGGATGG 1620  
QY 1621 TGATTTTTCGAGATCAAGATGTCGCAAAATATGCTGCCAGACTTGGTCAATCTTTTGG 1680  
Db 1621 TGATTTTTCGAGATCAAGATGTCGCAAAATATGCTGCCAGACTTGGTCAATCTTTTGG 1680  
QY 1681 TTCTCTCCAGAGACTTTTGAAGTGTCTTAGGATGAGATTGAAGTTATTCCTCGATGTA 1740  
Db 1681 TTCTCTCCAGAGACTTTTGAAGTGTCTTAGGATGAGATTGAAGTTATTCCTCGATGTA 1740  
QY 1741 GGTTCATGAAACAGCATATGCTTTTCTGATGGAATTTGTAATAATCTCTGTCATTTGC 1800  
Db 1741 GGTTCATGAAACAGCATATGCTTTTCTGATGGAATTTGTAATAATCTCTGTCATTTGC 1800  
QY 1801 TCATAGAGTTGCTCAAAATGTGGCTTCAATATACCCATCTCTTCCAGATTCGTTA 1860  
Db 1801 TCATAGAGTTGCTCAAAATGTGGCTTCAATATACCCATCTCTTCCAGATTCGTTA 1860  
QY 1861 TGGTGGATATAAGGTTTGTGGGTTGATCCGGATTCATCAATGAAGTGTCTTTGAG 1920  
Db 1861 TGGTGGATATAAGGTTTGTGGGTTGATCCGGATTCATCAATGAAGTGTCTTTGAG 1920  
QY 1921 AAAGACATGTCGAATATGATATCAGACAAATAAAGTTAGATGTCCTTGGATGGAGCA 1980  
Db 1921 AAAGACATGTCGAATATGATATCAGACAAATAAAGTTAGATGTCCTTGGATGGAGCA 1980

QY 1981 ATATCAGCCTTGTTATCTTAATCGTCAACTGATTAAGCTCTTGTCTACACTTGGAGTAA 2040  
DB 1981 ATATCAGCCTTGTTATCTTAATCGTCAACTGATTAAGCTCTTGTCTACACTTGGAGTAA 2040  
QY 2041 AGATGAAGTCTCGACAGAGCAAAAGAAAGCTGTAGATCAGCTTGATCTATCTTGA 2100  
DB 2041 AGATGAAGTCTCGACAGAGCAAAAGAAAGCTGTAGATCAGCTTGATCTATCTTGA 2100  
QY 2101 TGATCTTTGAAGGCACAGAGGCTTTGGAATGTGTCTCTCGAGAGAAACACTAAATAT 2160  
DB 2101 TGATCTTTGAAGGCACAGAGGCTTTGGAATGTGTCTCTCGAGAGAAACACTAAATAT 2160  
QY 2161 TCTCAGAGCAATGCTAAACTGTGGTTTATAAGCCTGTAGCTGAGCCCTTTCTTTCAATGAT 2220  
DB 2161 TCTCAGAGCAATGCTAAACTGTGGTTTATAAGCCTGTAGCTGAGCCCTTTCTTTCAATGAT 2220  
QY 2221 GTTGCAAACTCTCCGCGCATCAAGTGTCTCGATTTGCGGACTAGATCAAGAATATTTAT 2280  
DB 2221 GTTGCAAACTCTCCGCGCATCAAGTGTCTCGATTTGCGGACTAGATCAAGAATATTTAT 2280  
QY 2281 TCCAAATGGAAGAACAAATGATGGGATGTTTGGATGAATCCAGAACCTTGGAAATATGGTCA 2340  
DB 2281 TCCAAATGGAAGAACAAATGATGGGATGTTTGGATGAATCCAGAACCTTGGAAATATGGTCA 2340  
QY 2341 GGTGTTTGTTCAGTTTACTGTTGCTGACATGGAGAGTGTTCGACGATTTACATCCATT 2400  
DB 2341 GGTGTTTGTTCAGTTTACTGTTGCTGACATGGAGAGTGTTCGACGATTTACATCCATT 2400  
QY 2401 TAATACAGCAGATCCACCAACAGTAATTTCAATTCGAAAGGAAATGTGGTTGTTGCAAA 2460  
DB 2401 TAATACAGCAGATCCACCAACAGTAATTTCAATTCGAAAGGAAATGTGGTTGTTGCAAA 2460  
QY 2461 AATCCATGCTTCGATCCTCGTGGTGAATTCGTGTTTAAAGGCTGTAATGTTCGAGCGT 2520  
DB 2461 AATCCATGCTTCGATCCTCGTGGTGAATTCGTGTTTAAAGGCTGTAATGTTCGAGCGT 2520  
QY 2521 GCACCATGTTGATGTTGTTGTTATTCCTCAGAAAGGAAAGAACCTTCATCCGAATGA 2580  
DB 2521 GCACCATGTTGATGTTGTTGTTATTCCTCAGAAAGGAAAGAACCTTCATCCGAATGA 2580  
QY 2581 ATGTTCTGGAGAGTATGGATGGGATATCTACTTTGTTGCTGGGATCAAGACATGAT 2640  
DB 2581 ATGTTCTGGAGAGTATGGATGGGATATCTACTTTGTTGCTGGGATCAAGACATGAT 2640  
QY 2641 CCGGCCAAGCAAGTCCAGCCGATGAATATCTCCAGCAGCAGCAGCATACAGTTGGACCA 2700  
DB 2641 CCGGCCAAGCAAGTCCAGCCGATGAATATCTCCAGCAGCAGCAGCATACAGTTGGACCA 2700  
QY 2701 TGAATGCACAAATGAGGAAGTTGAAGAGTACTTCAACCACTATATTTGGAATGACAGTTT 2760  
DB 2701 TGAATGCACAAATGAGGAAGTTGAAGAGTACTTCAACCACTATATTTGGAATGACAGTTT 2760  
QY 2761 GGAATCATAGCAAAATGCCCATGCTCTATTTGCAGACAGAACCTGATATGCCATGAG 2820  
DB 2761 GGAATCATAGCAAAATGCCCATGCTCTATTTGCAGACAGAACCTGATATGCCATGAG 2820  
QY 2821 TGATCCATCAAAAACTTGCTGAGCTCTTTTCAATGTCAGTGGACTTTCCAAAGACTGG 2880  
DB 2821 TGATCCATCAAAAACTTGCTGAGCTCTTTTCAATGTCAGTGGACTTTCCAAAGACTGG 2880  
QY 2881 TGTTCCTCGTGAATACCAATCTCAGTTGGGCTTAAAGAAATACCCAGACTTCCATGATGAA 2940  
DB 2881 TGTTCCTCGTGAATACCAATCTCAGTTGGGCTTAAAGAAATACCCAGACTTCCATGATGAA 2940  
QY 2941 GCGGCAAGACAGCAGCTATATCTCAGAAAGAGTTATTTGAAAGCTTTTCCAGAAAGTGA 3000  
DB 2941 GCGGCAAGACAGCAGCTATATCTCAGAAAGAGTTATTTGAAAGCTTTTCCAGAAAGTGA 3000  
QY 3001 GGCACAAAGCACCTCAGGCTAGCTCTATTCGGACCTTTCAAGAGAGTTTGAAGGAGATC 3060  
DB 3001 GGCACAAAGCACCTCAGGCTAGCTCTATTCGGACCTTTCAAGAGAGTTTGAAGGAGATC 3060

QY 3061 ATATGATGCTGATATGGAAGTTGATGATGATTTGAAGATTACATTGACGAAGCTTTTGACTA 3120  
DB 3061 ATATGATGCTGATATGGAAGTTGATGATGATTTGAAGATTACATTGACGAAGCTTTTGACTA 3120  
QY 3121 CAAAACTCAATATGACAAAGCTGGGTAAATTAATGACTACTATGSCATAAAACAGA 3180  
DB 3121 CAAAACTCAATATGACAAAGCTGGGTAAATTAATGACTACTATGSCATAAAACAGA 3180  
QY 3181 GGCTGAAATACCTTAGTGTGGCATTATGAAGGCATCAAAAACTTTTGACCCGAGAAAAGA 3240  
DB 3181 GGCTGAAATACCTTAGTGTGGCATTATGAAGGCATCAAAAACTTTTGACCCGAGAAAAGA 3240  
QY 3241 TGCTGAGGCCATTAGTGTGCTGAGGGCTTGAGGAGGAGGCAAGAGCTGGTTCAA 3300  
DB 3241 TGCTGAGGCCATTAGTGTGCTGAGGGCTTGAGGAGGAGGCAAGAGCTGGTTCAA 3300  
QY 3301 GAGGCGTAATGATATAGATGACATGTTACCAAAAGCTTCGGCTTGGTACACGTTTACATA 3360  
DB 3301 GAGGCGTAATGATATAGATGACATGTTACCAAAAGCTTCGGCTTGGTACACGTTTACATA 3360  
QY 3361 TCATCCTACATATTGGGGTTGCTACAAATCAGGGGTTGAAAAGAGCTCAITTCATTAGCTT 3420  
DB 3361 TCATCCTACATATTGGGGTTGCTACAAATCAGGGGTTGAAAAGAGCTCAITTCATTAGCTT 3420  
QY 3421 TCCCTGGTGTGTTATGACCAAGCTAATCCAGATTAAAGAGGACAAAGCACCTAACAGGCC 3480  
DB 3421 TCCCTGGTGTGTTATGACCAAGCTAATCCAGATTAAAGAGGACAAAGCACCTAACAGGCC 3480  
QY 3481 AGTCTCAACTGTGCTCTCAGGGCTCAACTGAGTACAGATTAGTGTGAAATGAGA 3540  
DB 3481 AGTCTCAACTGTGCTCTCAGGGCTCAACTGAGTACAGATTAGTGTGAAATGAGA 3540  
QY 3541 TTCCAGTCGAGCGTTAAGCTGATATATATATATATATAGGTGTGATCATATAAGAAAC 3600  
DB 3541 TTCCAGTCGAGCGTTAAGCTGATATATATATATATATAGGTGTGATCATATAAGAAAC 3600  
QY 3601 TGTATGCACTTTTGACTACCTTTTGTCTTTAAACTGCATGAAGCTGCACATATATGTC 3660  
DB 3601 TGTATGCACTTTTGACTACCTTTTGTCTTTAAACTGCATGAAGCTGCACATATATGTC 3660  
QY 3661 AGTACTCAAGAACAGATGTACAGCTAAGTAAATATATATGATGTGATTTGAGTTTCAATC 3720  
DB 3661 AGTACTCAAGAACAGATGTACAGCTAAGTAAATATATGATGTGATTTGAGTTTCAATC 3720  
QY 3721 TTTCTTCTAAA 3731  
DB 3721 TTTCTTCTAAA 3731  
RESULT 2  
LOCUS LERDRP 3731 bp mRNA linear PLN 18-DEC-1998  
DEFINITION L.esculentum mRNA for RNA-directed RNA polymerase.  
ACCESSION Y10403  
VERSION Y10403.1 GI:4038591  
KEYWORDS RDRP gene; RNA-directed RNA polymerase.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
REFERENCE 1 (bases 1 to 3731)  
AUTHORS Schiebel,W., Pelissier,T., Riedel,L., Thalmair,S., Schiebel,R.,  
Kempe,D., Lottspeich,F., Sanger,H.L. and Wassenecker,M.  
TITLE Isolation of an RNA-directed RNA polymerase-specific cDNA clone  
from tomato  
JOURNAL Plant Cell 10 (12), 2087-2101 (1998)  
MEDLINE 98055198  
PUBMED 9836747  
REFERENCE 2 (bases 1 to 3731)  
AUTHORS Wassenecker,M.  
TITLE Direct Submission



JOURNAL Submitted (08-JAN-1997) M. Wassenecker, Max-Planck-inst. fuer Biochemie, Viroidforschung, Am Klopferspitz 18a, Planegg-Martinsried, 82152, FRG

FEATURES

Source Location/Qualifiers

1..3731

/organism="Lycopodium obscurum"

/strain="Rutgers"

/db\_xref="taxon:4081"

194..3538

/gene="RdRp"

194..3538

/gene="RdRp"

/EC\_number="2.7.7.48"

/codon\_start=1

/product="RNA-directed RNA polymerase"

/protein\_id="CAA71421.1"

/db\_xref="GI:4038592"

/translation="MGKTIQVGFVPLLSAEVVKFLEKTYGTGVCALVEVKOSKGGG  
 RAFAKVQFADNTSADKIIILANNRLVFGSSYLKAWEMKTDIVOLRAYVDWMDGTLNP  
 GCOISDDKFAVIGSTSVSIOFGIGIKKPEFLSSGSADYKQLSVENWVYLHRPYG  
 QNAQFLLIQFGNPRYIKRLNSCSIFKETFDDQWVRITDFPSPWIGLSSSLCLQFR  
 RGVLPNFEESEFHAERENNTILOTGTFVFSQKSAVPNPVQPEGISYIKILFKI  
 SSLVHGCIIPGALNVPFRLVDPRRRNVACIEHALEKLYIKECQYDPVRLWTEQYD  
 GYLKQPPKPSITLDDGLVVRVLVTPCKVYFCGPEVNSVRLNYSIEDINFL  
 RVFDEEMKLYSTDLDPKASTGSGVRTNIYERILSLRKGFVIGDKFEFLAFSS  
 QLDRNSVMEFASRPLGTLANDIRAMGDFSOIKNAVYARLGQFSGSKFSLVLRHE  
 IEVTPDKVFGTSYVSDGIGIKSGDFARVASKGLOYTPSAFOIRYGVYGVGVND  
 PDSMKLSLRKSMKSVESDNIKLDVLGSKYQPCVNLRLTLISLGVKDEVLQKQ  
 KEAVDQDAILHDSLKAQALBELSPGENTNLKALNCGKPKPQREPLSLMQLTFRA  
 SKLGLTRSRIFPNRMTMGCLDESRTLEYGVQVFTGAGHGFESDDLDHFPNRS  
 STNSNFIKLVVAVKPNCLPGDILRVKLVNRAHLMHVDVFPQKGRKRPNECS  
 GSDLDGDIYFVCDQDIMPQVQPMYPPAPSIQLDHDTVEEVEEFTNYIVNDSL  
 GIITANHVVFADREPDMASDPCKLAELFSLADFDVTPAEIPSOILRPEKYEYD  
 DKPDKTSYISERVIGLPRKVDKAPQASSTATFTRDVARRSYDADMEVDGFEDYIDE  
 AFQKTYEYNKGLNMDYVGIKTEAILSGGIMKASKTDFRDKDAEALSVAVRALRE  
 ARAWFERNDIDMLPKASAWVHYVHTPYWGCYNQGLKRAHFTSFPCWYVDLIQIK  
 KDKARNPVLNLSLRAQLSHRLVLK"

BASE COUNT 1064 a 688 c 850 g 1149 t

ORIGIN

Query Match 98.6%; Score 3680; DB 8; Length 3731;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 3730; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAATATCTTACTTACCCAGGAGTGTACTATCATCCCTCCCTCAAGCTTTTGT 60  
 Db 1 GAATATCTTACTTACTTACCCAGGAGTGTACTATCATCCCTCCCTCAAGCTTTTGT 60  
 Qy 61 GTTGTGATAAATTTGGTGTGCTTCAGTTTCAGTCACTACTGCTGGGTATTTAT 120  
 Db 61 GTTGTGATAAATTTGGTGTGCTTCAGTTTCAGTCACTACTGCTGGGTATTTAT 120  
 Qy 121 TTTGCATAACTTCAGGGGTATTCAGTTGGTGTAGCATTTGAAGTCGAACGACTT 180  
 Db 121 TTTGCATAACTTCAGGGGTATTCAGTTGGTGTAGCATTTGAAGTCGAACGACTT 180  
 Qy 181 GGAATTTGGGTACATGGGAAGACAATTCAGTTTTCGGATTCCTCTCTCTGC 240  
 Db 181 GGAATTTGGGTACATGGGAAGACAATTCAGTTTTCGGATTCCTCTCTCTGC 240  
 Qy 241 GGAAGTGGTAAAGTCATCTTAGAGAAATATACAGATATGGAATGTATGTGCAATTGA 300  
 Db 241 GGAAGTGGTAAAGTCATCTTAGAGAAATATACAGATATGGAATGTATGTGCAATTGA 300  
 Qy 301 GGTATAACAGTCCAAAGAGGATCTAGAGCATTTGCCAAGTTCATTTGCCGACAACAT 360  
 Db 301 GGTATAACAGTCCAAAGAGGATCTAGAGCATTTGCCAAGTTCATTTGCCGACAACAT 360  
 Qy 361 AAGTGTGACAAATATCATCTTTGGCTAATAACAGGCTGTATTTGGCTCTCTCTTATT 420  
 Db 361 AAGTGTGACAAATATCATCTTTGGCTAATAACAGGCTGTATTTGGCTCTCTCTTATT 420  
 Qy 421 GAAGGCTTGGGAATGAAAACTGATATTGTCCAACTGCGGGCATATGTGGATCAGATGA 480

Db 421 CAAGGCTTGGGAATGAAAACTGATATTGTCCAACTGCGGGCATATGTGGATCAGATGA 480  
 Qy 481 TGGCATAACTTTGAATTTTCGGATGTGAGATATCAGATGACAAAGTTTGCAGTGTGGGAAG 540  
 Db 481 TGGCATAACTTTGAATTTTCGGATGTGAGATATCAGATGACAAAGTTTGCAGTGTGGGAAG 540  
 Qy 541 TACAGAAGTTTCAATTTCAATTTGGCATTTGGATTTGAAGAAATTTTTTTCTTTTATCTAG 600  
 Db 541 TACAGAAGTTTCAATTTCAATTTGGCATTTGGATTTGAAGAAATTTTTTTCTTTTATCTAG 600  
 Qy 601 TGGTTCAAGTGTACATATAAATTCAGCTTTTCATATGAAAAATATATGCGAGTGTGTGCTCA 660  
 Db 601 TGGTTCAAGTGTACATATAAATTCAGCTTTTCATATGAAAAATATATGCGAGTGTGTGCTCA 660  
 Qy 661 TCGTCCATATGGTCAAAATGCTCAGTTTTCCTCATACAGTTATTTGGTGTCTCTCGGAT 720  
 Db 661 TCGTCCATATGGTCAAAATGCTCAGTTTTCCTCATACAGTTATTTGGTGTCTCTCGGAT 720  
 Qy 721 CTATAAGAGACTTGAATACTCTCTTATAGCTTCTTTAAGAAATCTCTGATGATCAGTG 780  
 Db 721 CTATAAGAGACTTGAATACTCTCTTATAGCTTCTTTAAGAAATCTCTGATGATCAGTG 780  
 Qy 781 GGTGAGGACACAGATTTCCCTCCATCTTGGATAGGCTATCTCTAGCTTATGTTTGA 840  
 Db 781 GGTGAGGACACAGATTTCCCTCCATCTTGGATAGGCTATCTCTAGCTTATGTTTGA 840  
 Qy 841 GTTCCGTAGGGTGTTCGTCTTCCAAATTCGAGGAAAGTTTTTCCACTATGCGAAGG 900  
 Db 841 GTTCCGTAGGGTGTTCGTCTTCCAAATTCGAGGAAAGTTTTTCCACTATGCGAAGG 900  
 Qy 901 TGAATAACAATATTACTTTACAGACTGGTTTACCTTTTCGCTCTCAAAAAATCGGCTCT 960  
 Db 901 TGAATAACAATATTACTTTACAGACTGGTTTACCTTTTCGCTCTCAAAAAATCGGCTCT 960  
 Qy 961 GGTTCCTCAATGTCCAGCTCCGGAAGAAATTTCAATTCCTCAAGATTTTGTTCAAAAT 1020  
 Db 961 GGTTCCTCAATGTCCAGCTCCGGAAGAAATTTCAATTCCTCAAGATTTTGTTCAAAAT 1020  
 Qy 1021 TAGTTCCTTTGGTACAGATGGATGATACCTCGGCGCAGCATTAATATGCTACTTTTCGG 1080  
 Db 1021 TAGTTCCTTTGGTACAGATGGATGATACCTCGGCGCAGCATTAATATGCTACTTTTCGG 1080  
 Qy 1081 ATTAGTTGATCCCTCGAAGGAAATGTGGATGATGATGATGATGATGATGATGATGATGAT 1140  
 Db 1081 ATTAGTTGATCCCTCGAAGGAAATGTGGATGATGATGATGATGATGATGATGATGATGAT 1140  
 Qy 1141 CTATATAAAGAGTGTGTATGATCCCTGAGGTGGCTCACTGAGCAGTATGATGGGTA 1200  
 Db 1141 CTATATAAAGAGTGTGTATGATCCCTGAGGTGGCTCACTGAGCAGTATGATGGGTA 1200  
 Qy 1201 TCTCAAGGTTAGACCAACCTCCAAAATCTCGGTCCTACATTTAGATGATGGTGTGTA 1260  
 Db 1201 TCTCAAGGTTAGACCAACCTCCAAAATCTCGGTCCTACATTTAGATGATGGTGTGTA 1260  
 Qy 1261 TGTAAAGAGGTTCTAGTAAACACCATGCAAGTTTATTTTGGTTCAGAGAGTTAAATGT 1320  
 Db 1261 TGTAAAGAGGTTCTAGTAAACACCATGCAAGTTTATTTTGGTTCAGAGAGTTAAATGT 1320  
 Qy 1321 TTCCAATCGGGTTCTCCGCAATTTATTTCTGAAGACATAGATACTTTCTTCGTTCTCTTT 1380  
 Db 1321 TTCCAATCGGGTTCTCCGCAATTTATTTCTGAAGACATAGATACTTTCTTCGTTCTCTTT 1380  
 Qy 1381 TGTGTAGTGAAGAGTGGGAGAACTGATTTCTACAGACTTATTACCAAAAAGCAAGTACTGG 1440  
 Db 1381 TGTGTAGTGAAGTGGGAGAACTGATTTCTACAGACTTATTACCAAAAAGCAAGTACTGG 1440  
 Qy 1441 AAGTGTGTGACAGACAAACATCTATGAGAGGATCTTATCAACTCTCGGAAAGCTTTGT 1500  
 Db 1441 AAGTGTGTGACAGACAAACATCTATGAGAGGATCTTATCAACTCTCGGAAAGCTTTGT 1500  
 Qy 1501 AATTGGTGATAAAAAATTTGAATTTCTTCATTCAGCAGCAGTTCGCGGATAAATTC 1560

Db 1501 AATTGGTGATAAAAAATTGAATTTCTGCAATTTTCATCGAGCCAGTTCGGGGAATAATTC 1560  
QY 1561 AGTGTGATGTTTGCATCAAGACCTGGCTTACTGCAAAATGATATAAGAGCTTGGATGG 1620  
Db 1561 AGTGTGATGTTTGCATCAAGACCTGGCTTACTGCAAAATGATATAAGAGCTTGGATGG 1620  
QY 1621 TGATTTTTCGCAGATCAAGAAATGTCGCAAAATATGCTGCCAGACTTGGTCAATCTTTGG 1680  
Db 1621 TGATTTTTCGCAGATCAAGAAATGTCGCAAAATATGCTGCCAGACTTGGTCAATCTTTGG 1680  
QY 1681 TTCTCCAGAGAGACTTTGAGTGTCTTCTAGGATGAGATTGAAGTATATCCCGATGTAAA 1740  
Db 1681 TTCTCCAGAGAGACTTTGAGTGTCTTCTAGGATGAGATTGAAGTATATCCCGATGTAAA 1740  
QY 1741 GGTTCATGAACAGACTATGCTTTCTGATGAATGTTGAATATCTGGTGAATTTGC 1800  
Db 1741 GGTTCATGAACAGACTATGCTTTCTGATGAATGTTGAATATCTGGTGAATTTGC 1800  
QY 1801 TCATAGAGTTGGCTCAAAATGTCGCTTCAATATACCCCATCTGCTTCCAGATTCTGTTA 1860  
Db 1801 TCATAGAGTTGGCTCAAAATGTCGCTTCAATATACCCCATCTGCTTCCAGATTCTGTTA 1860  
QY 1861 TGTGTGATATAAGGTGTTTGGGTGTTGATCCGGATTCATCAATGAAGTGTCTTTGAG 1920  
Db 1861 TGTGTGATATAAGGTGTTTGGGTGTTGATCCGGATTCATCAATGAAGTGTCTTTGAG 1920  
QY 1921 AAAGACATGTCGAATATGAATCAGACACATAAAGTTAGATGCTCTGGATGGAGCA 1980  
Db 1921 AAAGACATGTCGAATATGAATCAGACACATAAAGTTAGATGCTCTGGATGGAGCA 1980  
QY 1981 ATATCAGCCTTGTTAATCTGTCGCTCACTGATTAAGCTCTTGTCTACACTTTGGAGTAA 2040  
Db 1981 ATATCAGCCTTGTTAATCTGTCGCTCACTGATTAAGCTCTTGTCTACACTTTGGAGTAA 2040  
QY 2041 AGATGAAGTTCGACACAGACAAAGAAAGCTGTAGATCAGCTTGATGCTATCTTGA 2100  
Db 2041 AGATGAAGTTCGACACAGACAAAGAAAGCTGTAGATCAGCTTGATGCTATCTTGA 2100  
QY 2101 TGATTTCTTGAAGGCACAGAGGCTTTGGAATTTGATGCTCTCGAGAGAACTAAATAT 2160  
Db 2101 TGATTTCTTGAAGGCACAGAGGCTTTGGAATTTGATGCTCTCGAGAGAACTAAATAT 2160  
QY 2161 TCTCAGGCAATGCTAACTGTTGTTTAAAGCTGATGCTGAGCCCTTCTTTCAATGAT 2220  
Db 2161 TCTCAGGCAATGCTAACTGTTGTTTAAAGCTGATGCTGAGCCCTTCTTTCAATGAT 2220  
QY 2221 GTTGCAAACTTCCGGGCATCAAGTTGCTCGATTTGCGGACTAGATCAAGAATATTTAT 2280  
Db 2221 GTTGCAAACTTCCGGGCATCAAGTTGCTCGATTTGCGGACTAGATCAAGAATATTTAT 2280  
QY 2281 TCCAAATGGAAGAACAAATGATGGGATTTGGATGATGATCCAGAACTTGGAAATTTGGTCA 2340  
Db 2281 TCCAAATGGAAGAACAAATGATGGGATTTGGATGATGATCCAGAACTTGGAAATTTGGTCA 2340  
QY 2341 GGTGTTTGTTCAGTTTACTGTTGCTGGACATGGAGATTTTCTGACGATTTACATCCATT 2400  
Db 2341 GGTGTTTGTTCAGTTTACTGTTGCTGGACATGGAGATTTTCTGACGATTTACATCCATT 2400  
QY 2401 TAATAACAGCAGATCCACCAACAGTAATTTCAATCTGGAAGGAAATGTGGTTTGGCAAA 2460  
Db 2401 TAATAACAGCAGATCCACCAACAGTAATTTCAATCTGGAAGGAAATGTGGTTTGGCAAA 2460  
QY 2461 AAATCCATGCTGCATCTGTTGATATCGTGTTTTAAAGGCTGTAATGTTTCGAGCGCT 2520  
Db 2461 AAATCCATGCTGCATCTGTTGATATCGTGTTTTAAAGGCTGTAATGTTTCGAGCGCT 2520  
QY 2521 GCACCATGTTGATGTTGTTGTTATTCCTCAGAAAGGAAAGACCTCATCCGAATGA 2580  
Db 2521 GCACCATGTTGATGTTGTTGTTATTCCTCAGAAAGGAAAGACCTCATCCGAATGA 2580  
QY 2581 ATGTTCTGGAGTGATTTGGATGGGATATCTACTTTGTTGCTGGGATCAAGATGAT 2640  
Db 2581 ATGTTCTGGAGTGATTTGGATGGGATATCTACTTTGTTGCTGGGATCAAGATGAT 2640

QY 2641 CCGGCCAAGGCAAGTCCAGCCGATGGAATATCTCTCAGCACCCAGCATACAGTTGGACCA 2700  
Db 2641 CCGGCCAAGGCAAGTCCAGCCGATGGAATATCTCTCAGCACCCAGCATACAGTTGGACCA 2700  
QY 2701 TGATGTCACAATTCAGGAAGTTGAAGTACTTCCACCAACTATATTTGTGAATGACAGTTT 2760  
Db 2701 TGATGTCACAATTCAGGAAGTTGAAGTACTTCCACCAACTATATTTGTGAATGACAGTTT 2760  
QY 2761 GGGATCATAGCAATGCCATGTCGTATTTGCGACAGAGAAGCTGATATGGCCATGAG 2820  
Db 2761 GGGATCATAGCAATGCCATGTCGTATTTGCGACAGAGAAGCTGATATGGCCATGAG 2820  
QY 2821 TGATCCATGCAAAAACCTTGTGAGCTCTTTCAATTGCACTGAGCTTCCAAAGACTGG 2880  
Db 2821 TGATCCATGCAAAAACCTTGTGAGCTCTTTCAATTGCACTGAGCTTCCAAAGACTGG 2880  
QY 2881 TGTTCGCCCTGAAATACCATCTCAGTTGCGCCCTTAAAGAAATACCCAGACTTCATGGATAA 2940  
Db 2881 TGTTCGCCCTGAAATACCATCTCAGTTGCGCCCTTAAAGAAATACCCAGACTTCATGGATAA 2940  
QY 2941 GCGGCAAGACACAGCTATATCTCAGAAAGAGTTATTGGAAGCTTTTCAGGAAAGTAA 3000  
Db 2941 GCGGCAAGACACAGCTATATCTCAGAAAGAGTTATTGGAAGCTTTTCAGGAAAGTAA 3000  
QY 3001 GGACAAAGCACCTCAGGCTAGCTCTATCGGACCTTCAAGAGATGTTGCAAGGAGATC 3060  
Db 3001 GGACAAAGCACCTCAGGCTAGCTCTATCGGACCTTCAAGAGATGTTGCAAGGAGATC 3060  
QY 3061 ATATGATGCTGATATGGAAGTTGATGATTTGAAGTTACATTTGCAAGAGCTTTTGACTA 3120  
Db 3061 ATATGATGCTGATATGGAAGTTGATGATTTGAAGTTACATTTGCAAGAGCTTTTGACTA 3120  
QY 3121 CAAAACTGAATATGACACACAGCTGGTAAATTTATGGACTACTATGCAATAAACAAGA 3180  
Db 3121 CAAAACTGAATATGACACACAGCTGGTAAATTTATGGACTACTATGCAATAAACAAGA 3180  
QY 3181 GGCTGAAATAGTTAGTGGTGCATTTATGAAGGCATCAAAAACCTTTTGACCGCAGAAAAGA 3240  
Db 3181 GGCTGAAATAGTTAGTGGTGCATTTATGAAGGCATCAAAAACCTTTTGACCGCAGAAAAGA 3240  
QY 3241 TGCTGAGGCCATAGTGTGCTGTGAGGCCCTTGAGGAGGAGGACAGAGCTGGTCAA 3300  
Db 3241 TGCTGAGGCCATAGTGTGCTGTGAGGCCCTTGAGGAGGAGGACAGAGCTGGTCAA 3300  
QY 3301 GAGCGTAATGATATAGATGACATGTTACCAAAGGCTTCGGCTTGGTACCACGTTACATA 3360  
Db 3301 GAGCGTAATGATATAGATGACATGTTACCAAAGGCTTCGGCTTGGTACCACGTTACATA 3360  
QY 3361 TCATCCTACATATTTGGGTTGCTACATTCAGGGGTTGAAAAGAGCTCATTTTCATGCTT 3420  
Db 3361 TCATCCTACATATTTGGGTTGCTACATTCAGGGGTTGAAAAGAGCTCATTTTCATGCTT 3420  
QY 3421 TCCCTGGTGTGTTTATGACACAGCTAAATCCAGATTAAAGAGGACAAAGCCTAACAGGCC 3480  
Db 3421 TCCCTGGTGTGTTTATGACACAGCTAAATCCAGATTAAAGAGGACAAAGCCTAACAGGCC 3480  
QY 3481 AGTTCTCAACTTGTCACTCTCAGGGCTCACTGAGTCACAGATTAGTGTGTAATGAGA 3540  
Db 3481 AGTTCTCAACTTGTCACTCTCAGGGCTCACTGAGTCACAGATTAGTGTGTAATGAGA 3540  
QY 3541 TTCCAGTCGAGCGTTAAGCTGATATATATAATAGTAAAGGCTGATCATATAAGAAAAC 3600  
Db 3541 TTCCAGTCGAGCGTTAAGCTGATATATAATAGTAAAGGCTGATCATATAAGAAAAC 3600  
QY 3601 TGTATGATGTTTGGACTACCTTTTGTCTTAAACCTGCAATGAAGCTGCAACATATATGC 3660  
Db 3601 TGTATGATGTTTGGACTACCTTTTGTCTTAAACCTGCAATGAAGCTGCAACATATATGC 3660  
QY 3661 AGTACTCTAAGAAAACAGATGTACAGCTAAGTAAATATGATGTATGTTGAGTTTTCATC 3720  
Db 3661 AGTACTCTAAGAAAACAGATGTACAGCTAAGTAAATATGATGTATGTTGAGTTTTCATC 3720

QY 3721 TTCTCTCTAA 3731  
|||||  
Db 3721 TTCTCTCTAA 3731  
|||||

RESULT 3  
NTA011576 3505 bp mRNA linear PLN 07-JAN-1999  
LOCUS Nicotiana tabacum mRNA for RNA-directed RNA polymerase.  
DEFINITION AJ011576  
ACCESSION AJ011576  
VERSION AJ011576.1 GI:4138281  
KEYWORDS drp gene; RNA-directed RNA polymerase.  
SOURCE common tobacco.  
ORGANISM Nicotiana tabacum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
1 (bases 1 to 3505)  
Wassenegger, M.  
Direct Submission  
Submitted (16-OCT-1998) Wassenegger M., Max-Planck-Institut of  
Biochemistry Viroidresearch, Am Klopferspitz 18A, 82152  
Martinsried, FRG  
2 (bases 1 to 3505)  
Schiebel, W., Pelissier, T., Riedel, L., Thalmair, S., Schiebel, R.,  
Kemper, D., Lottspeich, F., Saenger, H.L. and Wassenegger, M.  
Isolation of an RNA-directed RNA polymerase-specific cDNA clone  
from Tomato  
Unpublished

FEATURES  
Location/Qualifiers  
1..3505  
/organism="Nicotiana tabacum"  
/cultivar="Petit Havana SRI"  
/db\_xref="taxon:4097"  
/tissue\_type="leaves"  
20..3370  
/gene="RdRp"  
20..3370  
/gene="RdRp"  
/codon\_start=1  
/evidence=experimental  
/product="RNA-directed RNA polymerase"  
/protein\_id="CAA09697.1"  
/db\_xref="GI:4138282"

translation="MGKTIQVGFYLLSAAVKKPLENHTGNGVYVLEVKQSKGGR  
RAFAQVQFANNSAEFIIDLASKGLYVGSYLKAMKTDIVQPTVYEHMDGVTLNF  
GGQISDKFAVLGSDVSKFGIGLKKIYFFLSHASADYKKLQSYENWQVYLRHPYG  
QNAQFLLQLGAPRIYKRLSDSCYSFFKETPDQDOWRTDFTPSWIGLSSCLLEFR  
NGVQPNFSEFFYKESNMQFIQTGFTFSQKALVPIVHLEPEGLPYKILFKI  
SSILQHCPCFPALANFFQLVPRRNACIEHALEKLYLKECCYDPVRLWTQYD  
EYLKQRLPKSPIITLDDGLVYVRVVTPCVYFCGPEVNSVRLNYSIEDINFL  
RYSFVDEWEKITHSDLLPRASTGNGRTDIYERILSLRNGFIIGDKRFEFLAFSSS  
QLRDSVWKEASRPGLTANDIRTMGDFQIRNVAKYARLGQSGSRETLVSGRHE  
VEIPDVACSLGNTYIFSDGIGKISADFAHRTVKGLQVTPSSFTQRYGKYGVA  
VPYSSMKLSLRKMLKYESNNIKLDVLGWSKYQCYLNRLQVTLTLLSLGVIDDVLEQ  
KQNEAVDQLDALHDSLKAQALEMSGENTNLKMLNCGMPDAFFLSMLLOT  
RASKLDDLTRITFNGRTWMGCLDESRTLEYGVQVVFQSGAGROFFFEESHFPND  
SGSANDCTILGNVYVAKNCPHLPDRIVRADVDPALHMDVDCVVPQKGRHPNE  
CSGSDIDGDIYFVCDPDLIPROVQSMYTPAPTQLDHDVTIEVEVEFTNYIIND  
SGIGIANAHVADPEPDMASDRCKLAQLFSIAVDPEKTPGVPAEIPSQLRPKEYPD  
FMKPKPTYPEKYGKLFQVKNKTPQASSIATFTRDVARKSDSDMVDGFEDI  
DEAFYKSYDNKGLMDYIGIKTEAELISGLGIMKASFTDRRKRDAEAGVAVACLR  
KEARAFYKRSIDMDLAKASAWHYVTHYTYWGLYNEGLKRDHFIISFPCVYDOLIO  
IKKAKRKPVPHLSLGSQSLRKLVIW"

BASE COUNT 1020 a 622 c 785 g 1078 t  
ORIGIN

Query Match 2.5%; Score 94; DB 8; Length 3505;  
Best Local Similarity 100.0%; Pred. No. 2.7e-39;  
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1516 ATTTGAATTTCTGCAATTTTCATCGAGCCAGTTGCGGGATAAATTCAGTGTGGATGTTTGC 1575  
|||||

Db 1342 ATTTGAATTTCTGCAATTTTCATCGAGCCAGTTGCGGGATAAATTCAGTGTGGATGTTTGC 1401  
QY 1576 ATCAAGACCTGCGCTTACTGCAAAATGATATAAGA 1609  
|||||  
Db 1402 ATCAAGACCTGCGCTTACTGCAAAATGATATAAGA 1435  
|||||

RESULT 4  
PHY011979 4579 bp DNA linear PLN 07-JAN-1999  
LOCUS Petunia hybrida RdRp gene, partial.  
DEFINITION AJ011979  
ACCESSION AJ011979  
VERSION AJ011979.1 GI:4138342  
KEYWORDS RdRp gene; RNA-directed RNA polymerase.  
SOURCE Petunia x hybrida.  
ORGANISM Petunia x hybrida  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Petunia.  
1 (bases 1 to 4579)  
Wassenegger, M.  
Direct Submission  
Submitted (16-OCT-1998) Wassenegger M., Max-Planck-Institut of  
Biochemistry Viroidresearch, Am Klopferspitz 18A, 82152  
Martinsried, FRG  
2 (bases 1 to 4579)  
Schiebel, W., Pelissier, T., Riedel, L., Thalmair, S., Schiebel, R.,  
Kemper, D., Lottspeich, F., Saenger, H.L. and Wassenegger, M.  
Isolation of an RNA-directed RNA polymerase-specific cDNA clone  
from Tomato  
Unpublished

FEATURES  
Location/Qualifiers  
1..4579  
/organism="Petunia x hybrida"  
/db\_xref="taxon:4102"  
/tissue\_type="leaves"  
1..4579  
/gene="RdRp"  
join(<1..1451,2420..2570,3856..>4579)  
/gene="RdRp"  
/codon\_start=1  
/evidence=experimental  
/product="RNA-directed RNA polymerase"  
/protein\_id="CAA09896.1"  
/db\_xref="GI:4138343"

translation="CIEHALEKLYNLSKNCVYVPPVRLTKQYDEYLRKRLPKSPITL  
DDGLMYVRVLVTPCKYVFLGPEVNSVRLNYSIEDINFLRVSFVDEWEKIFSM  
LLPRANNGRTDIYERILSLONGLVIGAKKFEFLAFSSQLRDSVWKEKIFSM  
AYDIRTMGDFQIRNVAKYARLGQSGSRETLVSGRHEIKYIPDVEHGTMYVS  
DGIGISADFAHRAVAKKCGQHQHTPSAQIRIGGKGVAVDVPSPMKLSLRKSMKYE  
SNNTKLDVLGWSKYQCYLNRLQVTLTSTLGKDVFPKQNEAVALDAILHDSLKA  
QEAALDMSGENTNLKMLNCGYLPAEPFLSMMLQTFRASKLLDLTRITRIFPKG  
RSMGCLDESRTLEYGVQVFTGAGGRQFVEESHFNDSGSANYFLIKGNVYVAKN  
PCLHPGDIRVLKAVDPALHMDVDCVVPQKGRHPNEGSGSDIDGDIYFVCDPDL  
IPRQVQPMYTPAPSIOLDHDVTIEVEVEFTNYMDGLGVIANAHVYFADREPNN  
AMSDPCLQAQLFSIAVDFTGVPAPISQRLRPKEYPDMKPKPTYPEKYGKLV  
FREYKDTQTASSITTRDVARKSDIDLEVDGFDYIDEAFYKSYDNKGLMD  
YIGIKTEAELISGLGIMKASFTDRRDAEAGVAVACLRKEARAFYKRSIDMDLAK  
ASANYHAYHTPYGTCYNEGLKRDHFIISFPCVYD"

<1..1451  
/gene="RdRp"  
/number=1  
1452..2419  
/gene="RdRp"  
/number=1  
2420..2570  
/gene="RdRp"  
/number=2  
2571..3855  
/gene="RdRp"  
/number=2  
3856..>4579  
/gene="RdRp"

exon  
intron  
exon  
intron  
exon

```

BASE COUNT      1361 a      789 c      943 g      1484 t      2 others
ORIGIN

Query Match      2.0%; Score 73; DB 8; Length 4579;
Best Local Similarity 100.0%; Pred. No. 9e-28;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2295 CAATGATGGATGTTGGATGAATCCAGAACCTTGCAATATGTCAGGTGTTTGTTCAGT 2354
Db 1181 CAAATGATGGATGTTGGATGAATCCAGAACCTTGCAATATGTCAGGTGTTTGTTCAGT 1240

QY 2355 TTAATGCTGCTGG 2367
Db 1241 TTAATGCTGCTGG 1253

RESULT 5
ATH011977
LOCUS      1538 bp      DNA      linear      PLN 07-JAN-1999
DEFINITION Arabidopsis thaliana RDRP gene, partial.
ACCESSION  AJ011977
VERSION     AJ011977.1 GI:4127462
KEYWORDS   RDRP gene; RNA-directed RNA polymerase.
SOURCE     thale cress.
ORGANISM   Arabidopsis thaliana
REFERENCE  1 (bases 1 to 1538)
AUTHORS   Schiebel, W., Pellissier, T., Riedel, L., Thalmair, S., Schiebel, R.,
TITLE     Kempt, D., Lottspeich, F., Saenger, H. L. and Wassenecker, M.
JOURNAL   Isolation of an RNA-directed RNA polymerase-specific cDNA clone
          from Tomato
          Unpublished
FEATURES   source
            Location/Qualifiers
            1..1538
            /organism="Arabidopsis thaliana"
            /cultivar="Columbia"
            /db_xref="taxon:3702"
            /tissue_type="leaves"
            1..1538
            /genes="RDRP"
            /genes="RDRP"
            /codon_start=1
            /evidence=experimental
            /product="RNA-directed RNA polymerase"
            /protein_id="CAA09894.1"
            /db_xref="GI:4375829"
            /translation="IRAMMGDFDHIRNVAKYAAARLQSPSSSRETLNVRSDIEVDPD
            VETISLGTIVFSDGIGKISAEFAKRVKCGLTETPSAFOIRYGYGVVAVDPNS
            SKLSLRKSNKFESENTKLDVLANSKYQPCYMNQLITLLSTLGVTSVFEKKOREV
            VDRDLRLTHPEARHAGLMAFGENTNLIKLCGKPADPEPLSMQLNFRSKL
            LELRTRIFISGRSMGCLDETRELTVEYGVVQISDPMRGRFRFIITGPVVAKNP
            CLHSGVFLQAVNPALNMVDCVFPQKGLRPHNPGSCSDLDGDIYFVCDQELK
            PPTSPSPMTPEPTQILDHVTIEVEEYFANYIVNDSLGIANAHTAFADKEPLKA
            FSDPCIELAKKSTAVDFPKTGVAAVIPHQLVXKVEYDFMEKPKDPTYESKNVIGKLF
            REVKRAPTLISTKSTFLDVASKSYDKMEVDGFEYVDEAFYQKANYDFKLGMLMDY
            YGKTR"
BASE COUNT      442 a      279 c      381 g      436 t
ORIGIN

Query Match      0.9%; Score 35; DB 8; Length 1538;
Best Local Similarity 100.0%; Pred. No. 7.2e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2699 CATGATGTCACAATTCAGGAAGTTGAAGTACTT 2733
Db 1060 CATGATGTCACAATTCAGGAAGTTGAAGTACTT 1094

RESULT 6
I18797/c
LOCUS      906 bp      DNA      linear      PAT 07-OCT-1996
DEFINITION Sequence 1 from patent US 5498533.
ACCESSION  I18797
VERSION     I18797.1 GI:1599152
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 906)
AUTHORS   Poovaiah, B. W., Takezawa, D., Han, T.-J. and An, G. H.
TITLE     Control of growth and development of potato plants
JOURNAL   Patent: US 5498533-A 1 12-MAR-1996;
          Location/Qualifiers
          1..906
          /organism="unknown"
BASE COUNT      254 a      151 c      219 g      282 t
ORIGIN

Query Match      0.8%; Score 30; DB 6; Length 906;
Best Local Similarity 100.0%; Pred. No. 0.00042;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3687 TAACTACTAAATATGATGATGTTGAGTTT 3716
Db 47 TAACTACTAAATATGATGATGTTGAGTTT 18

RESULT 7
POTCAM/c
LOCUS      906 bp      mRNA      linear      PLN 15-AUG-1995
DEFINITION Solanum tuberosum clone PCM1 calmodulin mRNA, complete cds.
ACCESSION  J04559
VERSION     J04559.1 GI:169476
KEYWORDS   calmodulin.
SOURCE     Solanum tuberosum (strain Russet Burbank) cDNA to mRNA.
ORGANISM   Solanum tuberosum
REFERENCE  1 (bases 1 to 906)
AUTHORS   Jena, P. K., Reddy, A. S. and Poovaiah, B. W.
TITLE     Molecular cloning and sequencing of a cDNA for plant calmodulin:
          signal-induced changes in the expression of calmodulin
          Proc. Natl. Acad. Sci. U.S.A. 86 (10), 3644-3648 (1989)
JOURNAL   89264480
PUBMED    2726741
COMMENT   Draft entry and printed copy of sequence [1] kindly submitted by
          B.W.Poovaiah, 16-MAY-1989.
FEATURES   source
            Location/Qualifiers
            1..906
            /organism="Solanum tuberosum"
            /strain="Russet Burbank"
            /db_xref="taxon:4113"
            /clone="PCM1"
            /tissue_type="tuber tip"
            <1..906
            81..530
            /codon_start=1
            /product="calmodulin"
            /protein_id="AAA74405.1"
            /db_xref="GI:169477"
            /translation="MAEQLTEQIAEFKEAFSLFDKDGCGCITTKELGTVMRSILGNP
            TEAEQLDMISEADADQNGTIDFPFLNLMARKMDTDEEELKEAFKPYFDKQNGFIS
            AAELRHVNTNLGKLTDEVDMEIRDGDGQGVNDEEFVNMMLAK"
BASE COUNT      254 a      151 c      219 g      282 t
ORIGIN

```

Query Match 0.8%; Score 30; DB 8; Length 906;  
 Best Local Similarity 100.0%; Pred. No. 0.00042;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3687 TAAGTACTAATATGATGATTCGATTGAGTTT 3716  
 Db 47 TAAGTACTAATATGATGATTCGATTGAGTTT 18

RESULT 8  
 LOCUS ARI45905/c  
 DEFINITION Sequence 4 from patent US 6218142.  
 ACCESSION ARI45905  
 VERSION ARI45905.1 GI:15109094  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 25)  
 AUTHORS Wassenegeger,M., Riedel,L., Schiebel,W. and Sanger,H.L.  
 TITLE Nucleic acid molecules encoding polypeptides having the enzymatic activity of an RNA-directed RNA polymerase (RDRP)  
 JOURNAL Patent: US 6218142-A 4 17-APR-2001;  
 FEATURES Location/Qualifiers  
 source 1..25  
 BASE COUNT 10 a 4 c 7 g 4 t  
 ORIGIN

Query Match 0.7%; Score 25; DB 6; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.29;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1838 CCATCTGCTTCAGATTTCGTTATG 1862  
 Db 25 CCATCTGCTTCAGATTTCGTTATG 1

RESULT 9  
 LOCUS ARI45907/c  
 DEFINITION Sequence 6 from patent US 6218142.  
 ACCESSION ARI45907  
 VERSION ARI45907.1 GI:15109096  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 25)  
 AUTHORS Wassenegeger,M., Riedel,L., Schiebel,W. and Sanger,H.L.  
 TITLE Nucleic acid molecules encoding polypeptides having the enzymatic activity of an RNA-directed RNA polymerase (RDRP)  
 JOURNAL Patent: US 6218142-A 6 17-APR-2001;  
 FEATURES Location/Qualifiers  
 source 1..25  
 BASE COUNT 4 a 5 c 10 g 6 t  
 ORIGIN

Query Match 0.7%; Score 25; DB 6; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.29;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2659 GCGGATGGAATATCCTCCAGCACCC 2683  
 Db 25 GCGGATGGAATATCCTCCAGCACCC 1

RESULT 10  
 LOCUS ARI45906/c  
 DEFINITION Sequence 5 from patent US 6218142.  
 ACCESSION ARI45906  
 VERSION ARI45906.1 GI:15109095  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 25)  
 AUTHORS Wassenegeger,M., Riedel,L., Schiebel,W. and Sanger,H.L.  
 TITLE Nucleic acid molecules encoding polypeptides having the enzymatic activity of an RNA-directed RNA polymerase (RDRP)  
 JOURNAL Patent: US 6218142-A 5 17-APR-2001;  
 FEATURES Location/Qualifiers  
 source 1..25  
 BASE COUNT 9 a 9 c 4 g 3 t  
 ORIGIN

Query Match 0.6%; Score 24; DB 6; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1879 TGTGGTGTTTCATCCGGATTCATC 1902  
 Db 24 TGTGGTGTTTCATCCGGATTCATC 1

RESULT 11  
 LOCUS AC095524/c  
 DEFINITION Rattus norvegicus clone CH230-7M23, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
 54 unordered pieces.  
 ACCESSION AC095524  
 VERSION AC095524.3 GI:21716983  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 REFERENCE 1 (bases 1 to 112247)  
 AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayale,M., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,Z., Lichtarge,O., Lied,C., Liu,J., Liu,M., Loulseged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenko,S., Oguh,M., Okwuon,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Fu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,

Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tasey, J., Taylor, C., Taylor, R., Telford, B., Thomas, R., Thomas, S.,  
 Umani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G. and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 112247)  
 Worley, K.C.  
 Direct Submission  
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 112247)  
 Worley, K.C.  
 Direct Submission  
 Submitted (10-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 9, 2002 this sequence version replaced gi:17942040.

# REFERENCE AUTHORS JOURNAL

## COMMENT

----- Genome Center of Medicine  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: SCPM  
 Center clone name: CH230-7M23  
 ----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 61250 bases at least Q40  
 Consensus quality: 65255 bases at least Q30  
 Consensus quality: 68415 bases at least Q20

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 54 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1248: contig of 1248 bp in length  
 1249 1348: gap of unknown length  
 1349 2739: contig of 1391 bp in length  
 2740 2839: gap of unknown length  
 2840 4110: contig of 1271 bp in length  
 4111 4210: gap of unknown length  
 4211 5710: contig of 1500 bp in length  
 5711 5810: gap of unknown length  
 5811 7102: contig of 1292 bp in length  
 7103 7202: gap of unknown length  
 7203 8664: contig of 1662 bp in length  
 8665 8964: gap of unknown length  
 8965 10025: contig of 1061 bp in length  
 10026 10125: gap of unknown length  
 10126 11842: contig of 1717 bp in length  
 11843 11942: gap of unknown length  
 11943 13271: contig of 1329 bp in length  
 13272 13371: gap of unknown length  
 13372 14409: contig of 1038 bp in length  
 14410 14509: gap of unknown length  
 14510 15797: contig of 1288 bp in length  
 15798 15897: gap of unknown length  
 15898 16919: contig of 1022 bp in length  
 16920 17019: gap of unknown length

17020 18674: contig of 1655 bp in length  
 18675 18774: gap of unknown length  
 18775 20074: contig of 1300 bp in length  
 20075 21780: contig of 1606 bp in length  
 21781 21880: gap of unknown length  
 21881 23609: contig of 1729 bp in length  
 23610 23709: gap of unknown length  
 23710 25341: contig of 1632 bp in length  
 25342 25441: gap of unknown length  
 25442 26734: contig of 1293 bp in length  
 26735 26834: gap of unknown length  
 26835 28072: contig of 1238 bp in length  
 28073 28172: gap of unknown length  
 28173 29239: contig of 1067 bp in length  
 29240 29339: gap of unknown length  
 29340 30668: contig of 1329 bp in length  
 30669 30768: gap of unknown length  
 30769 31788: contig of 1020 bp in length  
 31789 31888: gap of unknown length  
 31889 33345: contig of 1457 bp in length  
 33346 33445: gap of unknown length  
 33446 34447: contig of 1002 bp in length  
 34448 34547: gap of unknown length  
 34548 36750: contig of 2203 bp in length  
 36751 38214: contig of 1364 bp in length  
 38215 38314: gap of unknown length  
 38315 39423: contig of 1109 bp in length  
 39424 39523: gap of unknown length  
 39524 42257: contig of 2734 bp in length  
 42258 42357: gap of unknown length  
 42358 43881: contig of 1524 bp in length  
 43882 43981: gap of unknown length  
 43982 46294: contig of 2313 bp in length  
 46295 46394: gap of unknown length  
 46395 47523: contig of 1358 bp in length  
 47524 47853: gap of unknown length  
 47854 49218: contig of 1366 bp in length  
 49219 49318: gap of unknown length  
 49319 51384: contig of 2066 bp in length  
 51385 51484: gap of unknown length  
 51485 53269: contig of 1785 bp in length  
 53270 53693: gap of unknown length  
 53694 55675: contig of 2306 bp in length  
 55676 55775: gap of unknown length  
 55776 57592: contig of 1817 bp in length  
 57593 57692: gap of unknown length  
 57693 59528: contig of 1836 bp in length  
 59529 59628: gap of unknown length  
 59629 61385: contig of 1758 bp in length  
 61386 61485: gap of unknown length  
 61486 63514: contig of 2028 bp in length  
 63515 63614: gap of unknown length  
 63615 66113: contig of 2499 bp in length  
 66114 66213: gap of unknown length  
 66214 68443: contig of 2530 bp in length  
 68444 72358: gap of unknown length  
 72359 72458: contig of 3515 bp in length  
 72459 75507: contig of 3049 bp in length  
 75508 75607: gap of unknown length  
 75608 77932: contig of 2325 bp in length  
 77933 78032: gap of unknown length  
 78033 80521: contig of 2489 bp in length  
 80522 80621: gap of unknown length  
 80622 83316: contig of 2695 bp in length  
 83317 83416: gap of unknown length  
 83417 85700: contig of 2284 bp in length  
 85701 85800: gap of unknown length  
 85801 89181: contig of 3381 bp in length  
 89182 89281: gap of unknown length  
 89282 92357: contig of 3076 bp in length

[illegible]

/translation="MDDDFRNWGYEPAATKFNGLGLQMLQSLDRNTKPLPGRDP  
NLMGPNQSHQEPPIHMSYNWNOQDKDFNMLPVTATPNVGNLPSFAPSQ  
NMLHLLQTEENPVLDEEIVYQKRNKINAGSTPKAKPRPKDENSENNNNNT  
NVRVKKPAKSKVDLVINGSDGLPVICTCTGACQOQYCGGCGWQSACCTNIS  
MPLPMSTKRGARTSGRMSQGAFFKVKLEKLASDGFNGFNPIDLKSHWARHGTNKEV  
TIR"  
complement(13762..14781)  
/note="hypothetical protein"  
/codon\_start=1  
/evidence=not\_experimental  
/product="F10B6.6"  
/protein\_id="AAF79250.1"  
/db\_xref="GI:8778241"  
/translation="MKLWRAAVLVKQPSLIAADDLTAAVKATSHDELSIDTESA  
QPIYRHVLSPPSSKPLVSLISRVKTRPSAVALKGLMLMGGFLCKSTVAESIGRL  
PDLSSFGESNRINSKSGFNLFVAFALDRSLFHDGCRHNRNEESVILRLV  
LIRKQIIVLSLIRKIPIGENMIPVINEAMENVVSEIMEIYIGWICRHAEVLPNVRS  
KIGTEADLAKIVAKMSKGKELKYPFCCKDLGVNAQAEIPNFVRIPEADVILHDE  
LVRTAMESSESAETEAETAEETAEETAEETAEETAEETAEETAEETAEETAE  
DLISF"  
16027..16533  
/note="hypothetical protein"  
/codon\_start=1  
/evidence=not\_experimental  
/product="F10B6.7"  
/protein\_id="AAF79220.1"  
/db\_xref="GI:8778211"  
/translation="MSTCVYRECMRHAALGVYADGCREYSQPSGTGLCVACGCH  
PSYHRIDVISPQINHTFPTSLRVKQLARLKWKAETAEERNEEDDEETESTEK  
MYVQRKSKTAQREAMKYAAKWLKADRALREIRVFCGGVIRYHFKTW  
NNKRFYH"  
complement(join(17635..17815,17901..18019,18141..18194,  
18300..18406,18547..18611,18742..18950))  
/note="hypothetical protein"  
/codon\_start=1  
/evidence=not\_experimental  
/product="F10B6.8"  
/protein\_id="AAF79249.1"  
/db\_xref="GI:8778240"  
/translation="MEKFAFYLWFPFRLQCEPVIWFLFCQMILCKKORIFLFW  
YPELHLQOPLTKMVIWFLRIRWFDKOVCKSSSTKGLSVSGVAGALQGFNE  
IVRPVKSLGSKPKSLCPKSGKVLFGQALKVGTSLTFVGEAVRQKAGNLMQKSK  
EQSLVFESESFEMVNNKMSQSELGVILAKIFGSTAVAIAYVGYDYARKVLPFV  
WKKDLDSSQNGSSSEDFDISDF"  
complement(join(20071..20409,20499..20573,20648..20721,  
20800..20887,20957..21108,21188..21356,21448..21906,  
22006..22275,22367..22570,22840..22917))  
/note="hypothetical protein"  
/codon\_start=1  
/evidence=not\_experimental  
/product="F10B6.9"  
/protein\_id="AAF79248.1"  
/db\_xref="GI:8778239"  
/translation="MLEIESPTSLCFRTNTTCNALLRELQKIWVDIGESDAEKDRMLM  
ELEKCLPIYRVKDEANSKAQLHQSILVSTEARIASLLAALGVNESHSPKAKEGSK  
SLKEKLAVRNLEDLRLQDKRMKQVDIKRAQLEKSGEISGVSDQNLNMGSLAL  
DEQDLTLKNEYOTHLSRLOKESDRNLKVDLYNEVHTLCGVDFGQTVSEVRP  
SLHRDTHQSNISDDTLGLUHHMIHLKTERSVRFOKVDKHACNAILLFVDFHFA  
LAFPLSVQLQDVAGSLFELWMLDTSQEERTKFSYVYVRSESTITPILNLSST  
IEQVSAEYDFCNKLSKMLVNRRLTELENCLRAHIEADTISLEKSLIDSST  
KRLVDSPELLNIELHINKIEEARSKEIIDRIDRLWESACEENWLEEXNQDSTRYS  
AGRGHVNKHAERATVNVKIPSMVDNLIKKTLWEDETKSPFLYDGVRLVSLIEDY  
KLTRQOEERKERYDCKMODLLIKRESYIGSKPSRNSRVKRTNGYNGDASVPP  
TPRNSAGATNDIMTPRSYSSHRQNGYFKEVRLSTAPLNFVAIKESVSITYTSV  
CGSPDPSPLYN"  
join(26661..26948,27454..27610,27713..27814,27911..28038,  
28342..28470,28565..28632,28747..28969)  
/note="similar to acid phosphatase type 5 emb|CAB63938.1;  
similar to EST gb|N96048.1"  
/codon\_start=1  
/evidence=not\_experimental  
/product="F10B6.10"  
/protein\_id="AAF79221.1"

/db\_xref="GI:8778212"  
/translation="MYIYRDYTKITTKSTIPFLFFCFBSNLSMALLKHKPVLNVFY  
VYNLIIFSHSSTAELRLLQPSKTDGTVSFLVIGDWGRGSYNQVALQMGGEIGE  
Query Match 0.6% Score 23; DB 8; Length 132699;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2576 AATGAATGTTCTGGAGTGATTT 2598  
|||||  
Db 63331 AATGAATGTTCTGGAGTGATTT 63309  
|||||  
RESULT 13  
AC113586/204160 bp DNA linear HTG 26-JUN-2002  
LOCUS Mus musculus clone RP23-17D24, WORKING DRAFT SEQUENCE, 12 ordered  
pieces.  
AC113586 GI:21592013  
VERSION HTG: HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 204160)  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
Cooke, P., DeArellano, K., Dewar, K., Diaz, J., Dodge, S., Faro, S.,  
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Gande, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,  
Landers, T., Lehotzky, J., Levine, R., Liu, G., Maclean, C.,  
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,  
McEwan, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S.,  
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
TITLE Submitted (03-NAR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE 3 (bases 1 to 204160)  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,  
Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,  
Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J., Dodge, S.,  
Faro, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,  
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,  
Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Lindblad-Toh, K.,  
Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N.,  
Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,  
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,  
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,



Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Totham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Vial, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL

Direct Submission  
Submitted (26-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 26, 2002 this sequence version replaced gi:19881849.

COMMENT

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L13578

Center clone name: L7\_D\_24

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 200742 bases at least Q40

Consensus quality: 202113 bases at least Q30

Consensus quality: 202610 bases at least Q20

Insert size: 203000; agarose-1p

Insert size: 203060; sum-of-contigs

Quality coverage: 7.6 in Q20 bases; agarose-1p

Quality coverage: 7.6 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 12 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

\* 1 22703: contig of 22703 bp in length

\* 22704 22803: gap of 100 bp

\* 22804 23139: contig of 336 bp in length

\* 23140 23239: gap of 100 bp

\* 23240 24461: contig of 1222 bp in length

\* 24462 24561: gap of 100 bp

\* 24562 25820: contig of 1259 bp in length

\* 25821 25920: gap of 100 bp

\* 25921 28586: contig of 2666 bp in length

\* 28587 28686: gap of 100 bp

\* 28687 33028: contig of 4342 bp in length

\* 33029 33128: gap of 100 bp

\* 33129 52619: contig of 19491 bp in length

\* 52620 52719: gap of 100 bp

\* 52720 68399: contig of 15680 bp in length

\* 68400 68499: gap of 100 bp

\* 68500 95071: contig of 26572 bp in length

\* 95072 95171: gap of 100 bp

\* 95172 117191: contig of 22020 bp in length

\* 117192 117291: gap of 100 bp

\* 117292 153491: contig of 38200 bp in length

\* 153492 155591: gap of 100 bp

\* 155592 204160: contig of 48569 bp in length.

\* Location/Qualifiers

\* 1 .204160

\* /organism="Mus musculus"

\* /db\_xref="taxon:10090"

\* /clone\_lib="RP23-17D24"

\* /clone\_lib="RP23-17D24"

\* 1 .22703

\* /note="assembly\_fragment"

\* clone\_end:SP6

misc\_feature

vector\_side:left"  
22804..23139  
/note="assembly\_fragment"  
23240..24461  
/note="assembly\_fragment"  
24562..25820  
/note="assembly\_fragment"  
25921..28586  
/note="assembly\_fragment"  
28687..33028  
/note="assembly\_fragment"  
33129..52619  
/note="assembly\_fragment"  
52720..68399  
/note="assembly\_fragment"  
68500..95071  
/note="assembly\_fragment"  
95172..117191  
/note="assembly\_fragment"  
117292..153491  
/note="assembly\_fragment"  
153492..204160  
/note="assembly\_fragment"  
vector\_side:right"  
clone\_end:T7

BASE COUNT 59370 a 40877 c 39993 g 62817 t 1103 others  
ORIGIN

Query Match 0.6%; Score 22; DB 2; Length 204160;  
Best Local Similarity 100.0%; Pred. No. 7.4;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 673 TCAAAATGCTCAGTTTCTCTC 694  
|||||

Db 11543 TCAAAATGCTCAGTTTCTCTC 11522  
|||||

RESULT 14

AL732358.3

WPCOMMENT

Sequence split into 4 fragments LOCUS AL732358 Accession AL732358

Fragment Name	Begin	End
AL732358_0	1	110000
AL732358_1	100001	210000
AL732358_2	200001	310000
AL732358_3	300001	354683

Continuation (4 of 4) of AL732358 from base 300001 (AL732358 Homo sapiens chromosome

Query Match 0.6%; Score 21; DB 2; Length 54683;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 576 AGAAATTTTTTTCTTTTAT 596  
|||||

Db 33471 AGAAATTTTTTTCTTTTAT 33491  
|||||

RESULT 15

AL103999

LOCUS

DEFINITION

SEQUENCE SAMPLING.

AC103999

AC103999

VERSION

HTG; HTGS\_PHASE0.

KEYWORDS

ORGANISM

Homo sapiens

Homo sapiens

Mammalia; Eutheraia; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 62512)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 11, clone RP11-96012

Unpublished

AC103999 62512 bp DNA linear HTG 01-DEC-2001  
Homo sapiens chromosome 11 clone RP11-96012 map 11, LOW-PASS  
SEQUENCE SAMPLING.

AC103999

AC103999

HTG; HTGS\_PHASE0.

Homo sapiens

Homo sapiens

Mammalia; Eutheraia; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 62512)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 11, clone RP11-96012

Unpublished

REFERENCE  
AUTHORS

2 (bases 1 to 62512)  
 Birren, S., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckghalter, B.,  
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
 Choepel, Y., Collangelo, M., Collins, S., Collumore, A., Cook, A.,  
 Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,  
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
 Jones, C., Kamat, A., Karatas, A., Kells, C., Lakocque, K.,  
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
 MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
 McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,  
 Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Navlor, J., Nguyen, C.,  
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D.,  
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL

Direct Submission  
 Submitted (01-DEC-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

All repeats were identified using RepeatMasker:  
 Smit, A. P. A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WtBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center Project name: L21922

Center clone name: 96\_O\_12

-----  
 \* NOTE: This record contains 82 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1  
 674 773: contig of 673 bp in length  
 774 773: gap of 100 bp  
 1448 1448: contig of 675 bp in length  
 1449 1548: gap of 100 bp  
 1549 2177: contig of 629 bp in length  
 2178 2277: gap of 100 bp  
 2278 2837: contig of 610 bp in length  
 2838 2987: gap of 100 bp  
 2988 3684: contig of 677 bp in length  
 3685 3764: gap of 100 bp  
 3765 4423: contig of 665 bp in length  
 4430 4523: gap of 100 bp  
 4530 5195: contig of 666 bp in length  
 5196 5295: gap of 100 bp  
 5296 5980: contig of 685 bp in length  
 5981 6080: gap of 100 bp  
 6081 6751: contig of 671 bp in length  
 6752 6851: gap of 100 bp  
 6852 7512: contig of 661 bp in length  
 7513 7612: gap of 100 bp  
 7613 8283: contig of 671 bp in length  
 8284 8383: gap of 100 bp  
 8384 9047: contig of 664 bp in length  
 9048 9147: gap of 100 bp

9148 9771: contig of 624 bp in length  
 9772 9871: gap of 100 bp  
 10518 10518: contig of 647 bp in length  
 10519 10618: gap of 100 bp  
 10619 11261: contig of 643 bp in length  
 11262 11361: gap of 100 bp  
 11362 12014: contig of 653 bp in length  
 12015 12114: gap of 100 bp  
 12115 12792: contig of 678 bp in length  
 12793 12892: gap of 100 bp  
 12893 13563: contig of 671 bp in length  
 13564 13663: gap of 100 bp  
 13664 14328: contig of 665 bp in length  
 14329 14428: gap of 100 bp  
 14429 15098: contig of 670 bp in length  
 15099 15198: gap of 100 bp  
 15199 15867: contig of 669 bp in length  
 15868 15967: gap of 100 bp  
 15968 16644: contig of 677 bp in length  
 16645 16744: gap of 100 bp  
 16745 17393: contig of 649 bp in length  
 17394 17493: gap of 100 bp  
 17494 18166: contig of 673 bp in length  
 18167 18266: gap of 100 bp  
 18267 18921: contig of 655 bp in length  
 18922 19021: gap of 100 bp  
 19022 19693: contig of 672 bp in length  
 19694 19793: gap of 100 bp  
 19794 20457: contig of 664 bp in length  
 20458 20557: gap of 100 bp  
 20558 21240: contig of 683 bp in length  
 21241 21340: gap of 100 bp  
 21341 22011: contig of 671 bp in length  
 22012 22111: gap of 100 bp  
 22112 22790: contig of 679 bp in length  
 22791 22890: gap of 100 bp  
 22891 23565: contig of 675 bp in length  
 23566 23665: gap of 100 bp  
 23666 24352: contig of 687 bp in length  
 24353 24452: gap of 100 bp  
 24453 25116: contig of 664 bp in length  
 25117 25216: gap of 100 bp  
 25217 25879: contig of 663 bp in length  
 25880 25979: gap of 100 bp  
 25980 26623: contig of 644 bp in length  
 26624 26723: gap of 100 bp  
 26724 27377: contig of 654 bp in length  
 27378 27477: gap of 100 bp  
 27478 28147: contig of 670 bp in length  
 28148 28247: gap of 100 bp  
 28248 28906: contig of 659 bp in length  
 28907 29006: gap of 100 bp  
 29007 29680: contig of 674 bp in length  
 29681 29780: gap of 100 bp  
 29781 30444: contig of 664 bp in length  
 30445 30544: gap of 100 bp  
 30545 31204: contig of 660 bp in length  
 31205 31304: gap of 100 bp  
 31305 31981: contig of 677 bp in length  
 31982 32081: gap of 100 bp  
 32082 32758: contig of 677 bp in length  
 32759 32858: gap of 100 bp  
 32859 33518: contig of 660 bp in length  
 33519 33618: gap of 100 bp  
 33619 34274: contig of 656 bp in length  
 34275 34374: gap of 100 bp  
 34375 35024: contig of 650 bp in length  
 35025 35124: gap of 100 bp  
 35125 35783: contig of 659 bp in length  
 35784 35883: gap of 100 bp  
 35884 36563: contig of 680 bp in length  
 36564 36663: gap of 100 bp  
 36664 37338: contig of 675 bp in length



```

TITLE
JOURNAL
COMMENT
Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2000 this sequence version replaced gi:6910776.
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project information
Center project name: L3423
Center clone name: 14_M_16
-----
* NOTE: This record contains 88 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1
* 857: contig of 857 bp in length
* 958 957: gap of 100 bp
* 958 1852: contig of 895 bp in length
* 1853 1952: gap of 100 bp
* 1953 2811: contig of 859 bp in length
* 2812 2911: gap of 100 bp
* 2912 3773: contig of 862 bp in length
* 3774 3873: gap of 100 bp
* 3874 4737: contig of 864 bp in length
* 4738 4837: gap of 100 bp
* 4838 5710: contig of 873 bp in length
* 5711 5810: gap of 100 bp
* 5811 6692: contig of 882 bp in length
* 6693 6792: gap of 100 bp
* 6793 7651: contig of 859 bp in length
* 7652 7751: gap of 100 bp
* 7752 8643: contig of 892 bp in length
* 8644 8743: gap of 100 bp
* 8744 9620: contig of 877 bp in length
* 9621 9720: gap of 100 bp
* 9721 10603: contig of 883 bp in length
* 10604 10703: gap of 100 bp
* 10704 11582: contig of 879 bp in length
* 11583 11682: gap of 100 bp
* 11683 12567: contig of 885 bp in length
* 12568 12667: gap of 100 bp
* 12668 13534: contig of 867 bp in length
* 13535 13634: gap of 100 bp
* 13635 14479: contig of 845 bp in length
* 14480 14579: gap of 100 bp
* 14580 15449: contig of 870 bp in length
* 15450 15549: gap of 100 bp
* 15550 16432: contig of 883 bp in length
* 16433 16532: gap of 100 bp
* 16533 17413: contig of 881 bp in length
* 17414 17513: gap of 100 bp
* 17514 18415: contig of 902 bp in length
* 18416 18515: gap of 100 bp
* 18516 19408: contig of 893 bp in length
* 19409 19508: gap of 100 bp
* 19509 20387: contig of 879 bp in length
* 20388 20487: gap of 100 bp
* 20488 21390: contig of 909 bp in length
* 21397 21496: gap of 100 bp
* 21497 22335: contig of 839 bp in length
*
* 22336 22435: gap of 100 bp
* 22436 2295: contig of 860 bp in length
* 23296 23395: gap of 100 bp
* 23396 24296: contig of 901 bp in length
* 24297 24396: gap of 100 bp
* 24397 25267: contig of 871 bp in length
* 25268 25367: gap of 100 bp
* 25368 26243: contig of 876 bp in length
* 26244 26343: gap of 100 bp
* 26344 27175: contig of 832 bp in length
* 27176 27275: gap of 100 bp
* 27276 28150: contig of 875 bp in length
* 28151 28250: gap of 100 bp
* 28251 29158: contig of 908 bp in length
* 29159 29258: gap of 100 bp
* 29259 30151: contig of 893 bp in length
* 30152 30251: gap of 100 bp
* 30252 31148: contig of 897 bp in length
* 31149 31248: gap of 100 bp
* 31249 32137: contig of 889 bp in length
* 32138 32237: gap of 100 bp
* 32238 33086: contig of 849 bp in length
* 33087 33186: gap of 100 bp
* 33187 34115: contig of 929 bp in length
* 34116 34215: gap of 100 bp
* 34216 35090: contig of 875 bp in length
* 35091 35190: gap of 100 bp
* 35191 36057: contig of 867 bp in length
* 36058 36157: gap of 100 bp
* 36158 37002: contig of 845 bp in length
* 37003 37102: gap of 100 bp
* 37103 37953: contig of 851 bp in length
* 37954 38053: gap of 100 bp
* 38054 38945: contig of 892 bp in length
* 38946 39045: gap of 100 bp
* 39046 39946: contig of 901 bp in length
* 39947 40046: gap of 100 bp
* 40047 40936: contig of 890 bp in length
* 40937 41036: gap of 100 bp
* 41037 41905: contig of 869 bp in length
* 41906 42005: gap of 100 bp
* 42006 42906: contig of 901 bp in length
* 42907 43006: gap of 100 bp
* 43007 43885: contig of 879 bp in length
* 43886 43985: gap of 100 bp
* 43986 44824: contig of 839 bp in length
* 44825 44924: gap of 100 bp
* 44925 45812: contig of 888 bp in length
* 45813 45912: gap of 100 bp
* 45913 46787: contig of 875 bp in length
* 46788 46887: gap of 100 bp
* 46888 47800: contig of 913 bp in length
* 47801 47900: gap of 100 bp
* 47901 48795: contig of 895 bp in length
* 48796 48895: gap of 100 bp
* 48896 49767: contig of 872 bp in length
* 49768 49867: gap of 100 bp
* 49868 50813: contig of 946 bp in length
* 50814 50913: gap of 100 bp
* 50914 51818: contig of 903 bp in length
* 51819 51918: gap of 100 bp
* 51919 52792: contig of 874 bp in length
* 52793 52892: gap of 100 bp
* 52893 53779: contig of 887 bp in length
* 53780 53879: gap of 100 bp
* 53880 54764: contig of 885 bp in length
* 54765 54864: gap of 100 bp
* 54865 55785: contig of 921 bp in length
* 55786 55885: gap of 100 bp
* 55886 56772: contig of 887 bp in length
* 56773 56872: gap of 100 bp
* 56873 57820: contig of 948 bp in length
* 57821 57920: gap of 100 bp

```

```

* 57921 58797: contig of 877 bp in length
* 58798 58897: gap of 100 bp
* 58898 59751: contig of 854 bp in length
* 59752 59851: gap of 100 bp
* 59852 60731: contig of 880 bp in length
* 60732 60831: gap of 100 bp
* 60832 61723: contig of 892 bp in length
* 61724 61823: gap of 100 bp
* 61824 62748: contig of 925 bp in length
* 62749 62848: gap of 100 bp
* 62849 63738: contig of 890 bp in length
* 63739 63838: gap of 100 bp
* 63839 64717: contig of 879 bp in length
* 64718 64817: gap of 100 bp
* 64818 65674: contig of 857 bp in length
* 65675 65774: gap of 100 bp
* 65775 66735: contig of 961 bp in length
* 66736 66835: gap of 100 bp
* 66836 67703: contig of 868 bp in length
* 67704 67803: gap of 100 bp
* 67804 68689: contig of 886 bp in length
* 68690 68789: gap of 100 bp
* 68790 69656: contig of 867 bp in length
* 69657 69756: gap of 100 bp

Query Match          0.6%; Score 21; DB 2; Length 86478;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 579 AATTTTCTTTTCTTATCTA 599
|||||
Db 75142 AATTTTCTTTTCTTATCTA 75162

RESULT 18
AP004079
LOCUS
DEFINITION
Oryza sativa (japonica cultivar-group) chromosome 2 clone
OJ1067_B01, *** SEQUENCING IN PROGRESS ***
ACCESSION
AP004079.1 GI:15281358
VERSION
HTG; HTGS_PHASE2.
KEYWORDS
Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
clone:OJ1067_B01.
ORGANISM
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
1 Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC
clone:OJ1067_B01
Published Only in Database (2001)
JOURNAL
Sasaki,T., Matsumoto,T. and Yamamoto,K.
AUTHORS
TITLE
Direct Submission
Submitted (22-AUG-2001) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7458)
The nucleotide sequence of this BAC clone was generated by
combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
Location/Qualifiers
1..101706
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="2"
/clone="OJ1067_B01"
BASE COUNT 27699 a 22928 c 22675 g 28253 t 151 others
ORIGIN
Query Match          0.6%; Score 21; DB 2; Length 101706;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 579 AATTTTCTTTTCTTATCTA 599
|||||
Db 29464 AATTTTCTTTTCTTATCTA 29484

RESULT 19
AC099102
LOCUS
DEFINITION
Rattus norvegicus clone CH230-183A24, *** SEQUENCING IN PROGRESS
***, 54 unordered pieces.
ACCESSION
AC099102.4 GI:21729135
VERSION
HTG; HTGS_PHASE1.
KEYWORDS
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 122685)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.B., Amaratunga,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,J.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lied,C., Liu,J., Liu,W., Ioulseged,H.,
Lozard,R.J., Lu,X., Lucier,A., Lucier,K., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Minner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newtonson,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usman,K., Vasquez,L., Vera,V., Villalon,B., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (09-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
Worley, K.C.
3 (bases 1 to 122685)
Worley, K.C.
Direct Submission
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:20976259.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GGPI
Center clone name: CH230-183A24
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 63780 bases at least Q40
Consensus quality: 67858 bases at least Q30
Consensus quality: 71587 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 54 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
1
1183: contig of 1183 bp in length
1184: gap of unknown length
1284: contig of 1139 bp in length
2423: gap of unknown length
2523: contig of 1454 bp in length
3977: gap of unknown length
4077: contig of 1060 bp in length
5137: gap of unknown length
5237: contig of 1075 bp in length
6312: gap of unknown length
6412: contig of 1070 bp in length
7482: gap of unknown length
7582: contig of 1185 bp in length
8767: gap of unknown length
8867: contig of 1328 bp in length
10195: gap of unknown length
10295: contig of 1713 bp in length
12008: gap of unknown length
13317: contig of 1209 bp in length
13417: gap of unknown length
14501: contig of 1084 bp in length
14601: gap of unknown length
15882: contig of 1281 bp in length
15982: gap of unknown length
17715: contig of 1733 bp in length
17815: gap of unknown length
19609: contig of 1794 bp in length
19709: gap of unknown length
21281: contig of 1572 bp in length
21381: gap of unknown length
23129: contig of 1749 bp in length
23130: gap of unknown length
24676: contig of 1447 bp in length
24677: gap of unknown length
24777: contig of 1987 bp in length
26864: gap of unknown length
28515: contig of 1652 bp in length
28616: gap of unknown length
30279: contig of 1664 bp in length
30280: gap of unknown length
30380: contig of 1316 bp in length
31696: gap of unknown length
31796: contig of 1378 bp in length
33174: gap of unknown length
33273: contig of 1850 bp in length
35124: gap of unknown length
35223: contig of 1564 bp in length
36787: gap of unknown length
36887: gap of unknown length
36888: contig of 1375 bp in length
38463: gap of unknown length
38563: contig of 2499 bp in length
41062: gap of unknown length
41161: contig of 1747 bp in length
42908: gap of unknown length
43008: contig of 1895 bp in length
44904: gap of unknown length
45004: contig of 1416 bp in length
46420: gap of unknown length
46520: contig of 2105 bp in length
48624: gap of unknown length
48724: contig of 1272 bp in length
48937: gap of unknown length
50096: contig of 2352 bp in length
52648: gap of unknown length
52748: contig of 2051 bp in length
54799: gap of unknown length
54899: contig of 1865 bp in length
56764: gap of unknown length
56864: contig of 3179 bp in length
60043: gap of unknown length
60143: contig of 2160 bp in length
62403: gap of unknown length
62424: contig of 1821 bp in length
64324: gap of unknown length
65853: contig of 1529 bp in length
65953: gap of unknown length
69091: contig of 3138 bp in length
69191: gap of unknown length
71000: contig of 1809 bp in length
71100: gap of unknown length
73605: contig of 2505 bp in length
73705: gap of unknown length
76941: contig of 3236 bp in length
77041: gap of unknown length
81869: contig of 4828 bp in length
81969: gap of unknown length
84722: contig of 2753 bp in length
84822: gap of unknown length
87601: contig of 2779 bp in length
87701: gap of unknown length
90207: contig of 2506 bp in length
90307: gap of unknown length
94023: contig of 3716 bp in length
94123: gap of unknown length
97520: contig of 3397 bp in length
97620: gap of unknown length
97521: contig of 3828 bp in length
97621: gap of unknown length
101449: contig of 4144 bp in length
101549: gap of unknown length
105692: contig of 4541 bp in length
105792: gap of unknown length
110333: contig of 3203 bp in length
110433: gap of unknown length
113636: contig of 3413 bp in length
113736: gap of unknown length
117149: contig of 3413 bp in length

```

\* 117150 117249: gap of unknown length  
 \* 117250 122685: contig of 5436 bp in length.

Query Match 0.6%; Score 21; DB 2; Length 122685;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 574 GAAGAAATTTTCTTTT 594  
 |||||  
 Db 1935 GAAGAAATTTTCTTTT 1955

RESULT 20  
 HS960017/c  
 LOCUS  
 DEFINITION Human DNA sequence from clone 960017 on chromosome Xp11.21-11.22  
 Contains EST, CA repeat(DX9991), STS, GSS, complete sequence.

ACCESSION AL022166  
 VERSION AL022166.1 GI:4034471  
 KEYWORDS HTG; repeat polymorphism.  
 SOURCE Homo sapiens

ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 129226)

REFERENCE  
 AUTHORS Pavitt R.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-JAN-1999) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequests@sanger.ac.uk

COMMENT  
 On Dec 19, 1998 this sequence version replaced gi:4007559.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 only a small overlap as described above.

This sequence is the entire insert of clone 960017. The true right  
 end of clone 47K8 is at 100 in this sequence. This sequence has  
 been finished according to sequence map criteria as follows. An  
 attempt is made to resolve all sequencing problems, such as  
 compressions and repeats, but not necessarily within known  
 annotated human repeat sequence elements (e.g. Alu). Where the  
 feature is ambiguous, there is an annotation using the 'unsure'  
 feature key.

This sequence was generated from part of bacterial clone contigs of  
 human chromosome X, constructed by the Sanger Centre Chromosome X  
 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/ChrX>  
 960017 is from the library RPC15 constructed at the Roswell Park  
 Cancer Institute by the group of Pieter de Jong. For further  
 details see <http://bacpac.med.buffalo.edu/VECTOR:pcvPAC2>.

## FEATURES

Source  
 1. .129226  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="X"  
 /map="p11.21-11.22"  
 /clone="RP5-960017"  
 /clone\_lib="RPC1-5"  
 1. .3727  
 /note="L1PA6 repeat: matches 2414. .6143 of consensus"  
 3749. .3926  
 /note="HAL1 repeat: matches 343. .524 of consensus"  
 3796. .3994  
 /note="L1M4c repeat: matches 1497. .1709 of consensus"  
 complement(4521. .4899)  
 /note="match: GSS AQ024874"  
 4609. .4853  
 /note="L1ME3 repeat: matches 5910. .6160 of consensus"  
 4933. .5062  
 /note="HAL1 repeat: matches 685. .820 of consensus"  
 5491. .5783

/note="AluJb repeat: matches 1. .293 of consensus"  
 5491. .5581  
 /note="Single clone region"  
 5963. .6128  
 /note="L1M4 repeat: matches 2340. .2530 of consensus"  
 6199. .6397  
 /note="L1M4 repeat: matches 4632. .4843 of consensus"  
 6505. .6616  
 /note="L1M4 repeat: matches 2744. .2851 of consensus"  
 6617. .6877  
 /note="Alusq repeat: matches 39. .298 of consensus"  
 6878. .7031  
 /note="L1M4 repeat: matches 2851. .3019 of consensus"  
 7033. .7459  
 /note="L1M4 repeat: matches 4387. .4834 of consensus"  
 7583. .7888  
 /note="AluJo repeat: matches 1. .298 of consensus"  
 7890. .7934  
 /note="Alu repeat: matches 253. .297 of consensus"  
 7933. .8030  
 /note="L2 repeat: matches 2658. .2748 of consensus"  
 8084. .8359  
 /note="AluJb repeat: matches 1. .288 of consensus"  
 8558. .8776  
 /note="L1ME repeat: matches 5263. .5479 of consensus"  
 9202. .9215  
 /note="L1M4 repeat: matches 5423. .5437 of consensus"  
 9382. .9392  
 /note="L1M4 repeat: matches 5437. .5587 of consensus"  
 9623. .9927  
 /note="Alusq repeat: matches 1. .300 of consensus"  
 10984. .11279  
 /note="Alusx repeat: matches 1. .295 of consensus"  
 13148. .13430  
 /note="L1 repeat: matches 4789. .5073 of consensus"  
 13448. .14011  
 /note="L1MAL0 repeat: matches 5765. .6319 of consensus"  
 14021. .14527  
 /note="L1MA9 repeat: matches 5789. .6284 of consensus"  
 14528. .14863  
 /note="MER7A repeat: matches 1. .345 of consensus"  
 14864. .15772  
 /note="L1MA9 repeat: matches 4811. .5789 of consensus"  
 15774. .16392  
 /note="L1 repeat: matches 3938. .4556 of consensus"  
 16393. .16679  
 /note="L1PA8 repeat: matches 5886. .6163 of consensus"  
 16680. .17295  
 /note="L1 repeat: matches 3323. .3938 of consensus"  
 17298. .18190  
 /note="L1PA2 repeat: matches 5254. .6146 of consensus"  
 18218. .18855  
 /note="L1M1 repeat: matches 2659. .3329 of consensus"  
 18856. .19207  
 /note="L1MAL1 repeat: matches 1. .365 of consensus"  
 19208. .20181  
 /note="L1M1 repeat: matches 1776. .2659 of consensus"  
 20313. .20532  
 /note="Alusq/x repeat: matches 1. .303 of consensus"  
 20533. .20619  
 /note="L1M1 repeat: matches 1832. .1908 of consensus"  
 20620. .20912  
 /note="Alusx repeat: matches 1. .299 of consensus"  
 20913. .20929  
 /note="L1M1 repeat: matches 1815. .1832 of consensus"  
 20922. .21026  
 /note="L1M1 repeat: matches 1377. .1769 of consensus"  
 21063. .21414  
 /note="L1M1 repeat: matches 1398. .1716 of consensus"  
 21415. .21515  
 /note="HY3 repeat: matches 1. .101 of consensus"  
 21516. .21851  
 /note="L1M1 repeat: matches 1051. .1398 of consensus"

```

repeat_region 21855..22646
/note="LIPB3 repeat: matches 5358..6143 of consensus"
repeat_region 2467..22691
/note="MER44C repeat: matches 92..139 of consensus"
repeat_region 22692..23830
/note="LIPB3 repeat: matches 4162..5311 of consensus"
repeat_region 23821..24600
/note="LIM4C repeat: matches 1028..1815 of consensus"
repeat_region 24601..24890
/note="LIM4C repeat: matches 5995..6291 of consensus"
repeat_region 24891..24969
/note="LIM4C repeat: matches 952..1028 of consensus"
repeat_region 25001..25086
/note="LIMB6 repeat: matches 5860..5954 of consensus"
repeat_region 25094..26572
/note="LIPB1 repeat: matches 4687..5161 of consensus"
repeat_region 26560..27895
/note="LIMB6 repeat: matches 4549..5881 of consensus"
repeat_region 27921..28396
/note="LIMB8 repeat: matches 5764..6288 of consensus"
repeat_region 28397..30161
/note="LIM4 repeat: matches 2740..4562 of consensus"
repeat_region 30159..31561
/note="LIM4C repeat: matches 1044..2037 of consensus"
repeat_region 31575..35678
/note="LIPB13 repeat: matches 2494..6151 of consensus"
repeat_region 35679..35990
/note="LIMB8 repeat: matches 1..312 of consensus"
repeat_region 35991..38399
/note="LIPB13 repeat: matches 10..2494 of consensus"
repeat_region 38408..38560
/note="LIM4C repeat: matches 912..1066 of consensus"
repeat_region 38550..39328
/note="LIM4C repeat: matches 5..771 of consensus"
repeat_region 39793..39954
/note="MERA repeat: matches 10..189 of consensus"
repeat_region 41246..41985
/note="LIMB8 repeat: matches 5542..6288 of consensus"
repeat_region 42046..42183
/note="MLTIF repeat: matches 403..541 of consensus"
repeat_region 42188..42321
/note="LIMB6 repeat: matches 1..137 of consensus"
repeat_region 42413..42583
/note="MLTIF repeat: matches 268..431 of consensus"
repeat_region 42550..42850
/note="MLTIF repeat: matches 1..293 of consensus"
repeat_region 44044..44105
/note="31 copies 2 mer ac 68% conserved"
repeat_region 44195..44473
/note="MER69 repeat: matches 2195..2500 of consensus"
repeat_region 44503..45242
/note="MER69 repeat: matches 1..835 of consensus"
repeat_region 45528..50206
/note="LIMB8 repeat: matches 1468..6146 of consensus"
misc_feature 50628..50835
/note="match: STS L41742"
misc_feature complement(51196..51590)
/note="match: GSS AQ202670 clone R-58112"
repeat_region 52095..52406
/note="THEIC repeat: matches 1..371 of consensus"
repeat_region 53658..53962
/note="MLTIF repeat: matches 96..402 of consensus"
misc_feature complement(53950..54193)
/note="match: STS G03867"
repeat_region 54255..54513
/note="MLTIF repeat: matches 4..249 of consensus"
repeat_region 54610..55080
/note="LIMD2 repeat: matches 5914..6340 of consensus"

```

```

Query Match          0.6; Score 21; DB 9; Length 129226;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 576 AGAAATTTTCTTTTCTTTTAT 596
|||||TTTTTTTTTTTTTTTT|
Db 18223 AGAAATTTTCTTTTCTTTTAT 18203

RESULT 21
AC126008/c
LOCUS 131310 bp DNA linear HTG 21-AUG-2002
DEFINITION Medicago truncatula clone mth2-13e5, WORKING DRAFT SEQUENCE, 19
unordered pieces.
AC126008
VERSION AC126008.9 GI:22380755
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Euphorbiaceae; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
REFERENCE 1 (bases 1 to 131310)
AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,
Cook, D., Kim, D. and Roe, B.A.
TITLE Medicago truncatula BAC Clone mth2-13e5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 131310)
AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,
Cook, D., Kim, D. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-2002) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 131310)
AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,
Cook, D., Kim, D. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-2002) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Aug 21, 2002 this sequence version replaced gi:22325003.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2042: contig of 2042 bp in length
* 2043 2142: gap of unknown length
* 2143 4623: contig of 2481 bp in length
* 4624 4723: gap of unknown length
* 4724 6733: contig of 2010 bp in length
* 6734 6833: gap of unknown length
* 6834 8852: contig of 2019 bp in length
* 8853 8952: gap of unknown length
* 8953 11256: contig of 2304 bp in length
* 11257 11356: gap of unknown length
* 11357 14040: contig of 2684 bp in length
* 14041 14140: gap of unknown length
* 14141 16181: contig of 2041 bp in length
* 16182 16281: gap of unknown length
* 16282 18706: contig of 2425 bp in length
* 18707 18806: gap of unknown length
* 18807 22430: contig of 3624 bp in length
* 22431 22530: gap of unknown length
* 22531 28441: contig of 5911 bp in length
* 28442 28541: gap of unknown length
* 28542 34988: contig of 6447 bp in length

```



```

* 34989 35088: gap of unknown length
* 35089 42615: contig of 7527 bp in length
* 42616 42715: gap of unknown length
* 42716 51635: contig of 8920 bp in length
* 51636 51735: gap of unknown length
* 51736 59775: contig of 8240 bp in length
* 59776 60075: gap of unknown length
* 60076 66561: contig of 6486 bp in length
* 66562 80547: gap of unknown length
* 80548 80647: gap of 13886 bp in length
* 80648 98223: contig of 17576 bp in length
* 98224 98323: gap of unknown length
* 98324 114237: contig of 15914 bp in length
* 114238 114337: gap of unknown length
* 114338 131310: contig of 16973 bp in length.
FEATURES
    Location/Qualifiers
        1..131310
            /organism="Medicago truncatula"
            /db_xref="taxon:3880"
            /clone="mth2-13e5"
            /clone_lib="Medicago truncatula BAC library H2"
BASE COUNT 43027 a 22119 c 22734 g 41572 t 1858 others
ORIGIN
Query Match      0.6%; Score 21; DB 2; Length 131310;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 TGCATAACTTCAGGGGTATT 143
|||||
Db 29192 TGCATAACTTCAGGGGTATT 29172

RESULT 22
AC117166/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-387J16, *** SEQUENCING IN PROGRESS
***, 27 unordered pieces.
AC117166 135458 bp DNA linear HTG 18-JUL-2002
AC117166.2 GI:21746866
HTG: HGSC_PHASE1.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 135458)
Muzny,B.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayelle,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovicic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Luchterghe,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,

```

```

Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokkwo,S., Ogdu,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Saverly,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 135458)
Worley,K.C.
Direct Submission
Submitted (08-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 135458)
Worley,K.C.
Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:20069678.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GVDH
Center clone name: CH230-387J16
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 108275 bases at least Q40
Consensus quality: 113987 bases at least Q30
Consensus quality: 117413 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 27 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 1652: contig of 1652 bp in length
1752: gap of unknown length
1753 2894: contig of 1142 bp in length
2895 2995: gap of unknown length
2995 4031: contig of 1057 bp in length
4031 4151: gap of unknown length
4151 6042: contig of 1891 bp in length
6042 6143: gap of unknown length
6143 7419: contig of 1277 bp in length
7419 7519: gap of unknown length
7519 8797: contig of 1278 bp in length
8797 8898: gap of unknown length
8898 11467: contig of 2570 bp in length
11467 11568: gap of unknown length
11568 14223: contig of 2656 bp in length
14223 14324: gap of unknown length
14324 16190: contig of 1867 bp in length

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```



/note="18 copies 6 mer atat 67% conserved"  
11464..11569  
/note="53 copies 2 mer ta 68% conserved"  
11576..11703  
/note="64 copies 2 mer at 71% conserved"  
12246..12742  
/note="match: GSS: Em:AQ429690"  
12466..12761  
/note="AluJb repeat: matches 1..298 of consensus"  
13264..13725  
/note="match: GSS: Em:AQ412829"  
complement(13345..13537)  
/note="match: GSS: Em:AQ342536"  
13405..13443  
/note="HSMAR1 repeat: matches 1253..1287 of consensus"  
13474..13501  
/note="7 copies 4 mer attc 92% conserved"  
13844..13873  
/note="15 copies 2 mer ca 90% conserved"  
13878..14299  
/note="L2 repeat: matches 1578..2012 of consensus"  
14074..14548  
/note="match: GSS: Em:AQ080563"  
14470..14562  
/note="MLR1H repeat: matches 362..447 of consensus"  
14888..15000  
/note="L2 repeat: matches 2473..2586 of consensus"  
15001..15117  
/note="L1P4 repeat: matches 6030..6146 of consensus"  
15118..15251  
/note="L2 repeat: matches 2586..2733 of consensus"  
16315..16551  
/note="MIR repeat: matches 12..251 of consensus"  
17577..17862  
/note="match: GSS: Em:AQ066119"  
19153..19302  
/note="MIR repeat: matches 92..259 of consensus"  
complement(20295..20698)  
/note="match: GSS: Em:AQ080901"  
20832..21302  
/note="match: GSS: Em:AQ528326"  
20943..21095  
/note="MIR repeat: matches 88..245 of consensus"  
22090..22403  
/note="AluX repeat: matches 1..312 of consensus"  
22676..22785  
/note="MIR repeat: matches 148..260 of consensus"  
22786..23092  
/note="AluX repeat: matches 1..305 of consensus"  
23093..23229  
/note="MIR repeat: matches 21..148 of consensus"  
23832..24499  
/note="match: GSS: Em:AQ029286  
match: STS: Em:G51371"  
24764..25027  
/note="L1M2 repeat: matches 5792..6050 of consensus"  
25067..25131  
/note="L1P15 repeat: matches 6088..6152 of consensus"  
25132..25456  
/note="L1P16 repeat: matches 5701..6152 of consensus"  
25449..25539  
/note="L1M4 repeat: matches 5474..5563 of consensus"  
27676..27882  
/note="MLR1H repeat: matches 205..476 of consensus"  
28931..29139  
/note="MIR repeat: matches 40..262 of consensus"  
31511..31542  
/note="8 copies 4 mer aatg 90% conserved"  
31694..31790  
/note="MERA repeat: matches 88..184 of consensus"  
31805..31959  
/note="L2 repeat: matches 2548..2708 of consensus"  
complement(31985..32329)

/note="match: GSS: Em:AQ103099"  
32447..32919  
/note="match: GSS: Em:AQ785577"  
32523..33151  
/note="match: GSS: Em:AQ540822"  
32533..33003  
/note="match: GSS: Em:AQ534839"  
32964..33060  
/note="L2 repeat: matches 2607..2704 of consensus"  
33631..33980  
/note="MLR1H repeat: matches 140..547 of consensus"  
34288..34451  
/note="L1P4 repeat: matches 5981..6146 of consensus"  
34765..34955  
/note="MIR repeat: matches 1..236 of consensus"  
35237..35500  
/note="132 copies 2 mer aa 59% conserved"  
35251..35418  
/note="28 copies 6 mer agagag 64% conserved"  
35267..35446  
/note="45 copies 4 mer agaa 64% conserved"  
36398..36717  
/note="LRR1C repeat: matches 50..374 of consensus"  
36846..37128  
/note="match: GSS: Em:A2243634"  
37133..37394  
/note="match: GSS: Em:A2243634"  
37717..37899  
/note="MLR1J repeat: matches 352..535 of consensus"  
38110..38224  
/note="MLR1H repeat: matches 1..145 of consensus"  
38651..38690  
/note="20 copies 2 mer gt 100% conserved"  
39212..39848  
/note="match: GSS: Em:AQ355762"  
39250..39315  
Query Match 0.8%; Score 21; DB 9; Length 137723;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1941 AATCAGACACATAAAGTTAG 1961  
|||||  
Db 78002 AATCAGACACATAAAGTTAG 77982  
RESULT 24  
AP004023/c 138931 bp DNA linear HTG 21-MAR-2002  
LOCUS Oryza sativa (japonica cultivar-group) chromosome 2 clone  
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 2 clone  
OJ1126.D09, \*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.  
ACCESSION AP004023  
VERSION AP004023.1 GI:15130685  
KEYWORDS HTG; HTGS\_PHASE2.  
SOURCE Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,  
clone:OJ1126.D09.  
ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.  
REFERENCE  
1  
AUTHORS  
TITLE  
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC  
clone:OJ1126.D09  
JOURNAL  
TITLE  
Published Only in Database (2001)  
REFERENCE  
2 (bases 1 to 138931)  
AUTHORS  
TITLE  
Submitted (08-AUG-2001) Takuji Sasaki, National Institute of  
Agricultural Resources, Rice Genome Research Program; Kannondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,  
Tel:81-298-38-7441, Fax:81-298-38-7468)

COMMENT The nucleotide sequence of this BAC clone was generated by combining Monsanto and RGP-Japan sequencing data.  
 NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

FEATURES  
 source  
 1. .138931  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /chromosome="2"  
 /clone="OJ1126.D09"

BASE COUNT 39463 a 30336 c 30892 g 37874 t 566 others  
 ORIGIN

Query Match 0.6%; Score 21; DB 2; Length 138931;  
 Best Local Similarity 100.0%; Pred.No. 27;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 579 AATTTTTCCTTTTATCTA 599  
 |||||  
 Db 12861 AATTTTTCCTTTTATCTA 12841

RESULT 25  
 AC110620 142234 bp DNA linear PRI 05-JUN-2002  
 LOCUS Homo sapiens BAC clone RP11-683I7 from 2, complete sequence.  
 DEFINITION AC110620  
 ACCESSION AC110620  
 VERSION 3 GI:21326423  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 142234)  
 Sulston, J.E. and Waterston, R.  
 Toward a complete human genome sequence  
 Genome Res. 8 (11), 1097-1108 (1998)  
 99063792  
 9847074  
 2 (bases 1 to 142234)  
 Trani, L. and Kozlowitz, A.  
 The sequence of Homo sapiens BAC clone RP11-683I7  
 Unpublished (2001)  
 3 (bases 1 to 142234)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (13-FEB-2002) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 4 (bases 1 to 142234)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (19-FEB-2002) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 5 (bases 1 to 142234)  
 Waterston, R.  
 Direct Submission  
 Submitted (05-JUN-2002) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Jun 5, 2002 this sequence version replaced gi:18702445.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC

Web site: http://genome.wustl.edu/gsc  
 Contact: sapiens@wustl.wustl.edu  
 ----- Summary Statistics  
 Center project name: H\_NR0683I07  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

#### SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org  
 VECTOR: pBACE3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-702L18, 2000 bp overlap; the clone sequenced to the right is RP11-341H1, 2000 bp overlap. Actual start of this clone is at base position 107884 of RP11-702L18; actual end is at base position 33545 of RP11-341H1.

Polymorphisms have been identified between AC084367 and AC110620. Data from AC084367 was used to finish this clone, AC110620.

FEATURES	source
Location/Qualifiers	1. .142234
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-683I7"
	/clone_lib="RP11-11"
	1. .56
repeat_region	/rpt_family="Alu"
	142. .166
repeat_region	/rpt_family="AT-rich"
	176. .400
repeat_region	/rpt_family="L1"
	402. .830
repeat_region	/rpt_family="MER1_type"
	654. .987
repeat_region	/rpt_family="L1"
	1043. .1198
repeat_region	/rpt_family="L1"
	1322. .1432
repeat_region	/rpt_family="MER1_type"
	1437. .1578
repeat_region	/rpt_family="Tc2"
	1581. .1876
repeat_region	/rpt_family="Alu"
	2007. .2511
repeat_region	/rpt_family="L1"
	2874. .2975
repeat_region	/rpt_family="L1"

```

repeat_region 2976..3321 /rpt_family="Alu"
repeat_region 3573..3986 /rpt_family="L1"
repeat_region 3987..4299 /rpt_family="Alu"
repeat_region 4300..4724 /rpt_family="L1"
repeat_region 4725..5018 /rpt_family="Alu"
repeat_region 5019..5214 /rpt_family="L1"
repeat_region 5244..6154 /rpt_family="L1"
repeat_region 6235..6603 /rpt_family="L1"
repeat_region 6609..6785 /rpt_family="ERVL"
repeat_region 6812..7007 /rpt_family="ERVL"
repeat_region 7008..7196 /rpt_family="L1"
repeat_region 7244..7285 /rpt_family="MaLR"
repeat_region 7286..7583 /rpt_family="Alu"
repeat_region 7584..7854 /rpt_family="MaLR"
repeat_region 8885..9164 /rpt_family="L2"
repeat_region 10257..10545 /rpt_family="Alu"
repeat_region 10810..10837 /rpt_family="AT-rich"
repeat_region 10907..10933 /rpt_family="AT-rich"
repeat_region 10934..11245 /rpt_family="Alu"
repeat_region 11981..12615 /rpt_family="ERVL"
repeat_region 15756..15931 /rpt_family="MERL_type"
repeat_region 17805..17830 /rpt_family="AT-rich"
repeat_region 20488..21782 /rpt_family="G-rich"
misc_feature 20593..20621 /rpt_family="GC-rich"
repeat_region 20577..21026 /rpt_family="GC-rich"
repeat_region 21189..21244 /rpt_family="GC-rich"
repeat_region 21376..21471 /rpt_family="G-rich"
repeat_region 22393..22919 /rpt_family="ERVL"
repeat_region 23225..23321 /rpt_family="MIR"
repeat_region 23500..23629 /rpt_family="MERL_type"
repeat_region 23629..23656 /rpt_family="MERL_type"
repeat_region 23657..23949 /rpt_family="Alu"
repeat_region 23950..24016 /rpt_family="MERL_type"
repeat_region 24306..24743 /rpt_family="ERVL"
repeat_region 24762..24834 /rpt_family="MERL_type"
repeat_region 25123..25244 /rpt_family="MIR"
repeat_region 25288..25995 /rpt_family="MIR"

/rpt_family="MERL_type"
25996..26326 /rpt_family="Alu"
26327..26352 /rpt_family="MERL_type"
26353..26616 /rpt_family="L1"
26528..27528 /rpt_family="L1"
27530..27632 /rpt_family="Alu"
27633..27772 /rpt_family="L1"
28834..28948 /rpt_family="L1"

Query Match 0.58; Score 21; DB 9; Length 142234;
BestLocal Similarity 100.0%; Pred.No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 579 AATTTTTCCTTTTATCTA 599
|||||
DB 2029 AATTTTTCCTTTTATCTA 2049

RESULT 26
AC108260
LOCUS
DEFINITION AC108260 150055 bp DNA linear HTG 17-JUL-2002
Rattus norvegicus clone CH230-190K14, *** SEQUENCING IN PROGRESS
***, 69 unordered pieces.
ACCESSION AC108260
VERSION AC108260.3 GI:21743959
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 150055)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaraturge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbora,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.F.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeLaaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Kovah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissege,H.,
Lozato,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokkwo,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,

```

```

Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
2 (bases 1 to 150055)
Worley K.C.
Direct Submission
Submitted (27-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 150055)
Worley K.C.
Direct Submission
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 13, 2002 this sequence version replaced gi:18846337.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GM2I
Center clone name: CH230-190K14
----- Summary Statistics
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 80464 bases at least Q40
Consensus quality: 83462 bases at least Q30
Consensus quality: 85753 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 69 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1045: contig of 1045 bp in length
* 1046 1145: gap of unknown length
* 1146 2594: contig of 1449 bp in length
* 2595 2894: gap of unknown length
* 2895 3847: contig of 1153 bp in length
* 3848 3947: gap of unknown length
* 3948 5018: contig of 1071 bp in length
* 5019 5118: gap of unknown length
* 5119 6480: contig of 1362 bp in length
* 6481 6580: gap of unknown length
* 6581 7896: contig of 1116 bp in length
* 7897 7996: gap of unknown length
* 7997 9007: contig of 1211 bp in length
* 9008 9107: gap of unknown length
* 9108 10854: contig of 1747 bp in length
* 10855 10954: gap of unknown length
* 10955 11962: contig of 1008 bp in length
* 11963 12062: gap of unknown length
* 12063 13473: contig of 1411 bp in length
* 13474 13573: gap of unknown length
* 13574 15411: contig of 1838 bp in length
* 15412 15511: gap of unknown length
* 15512 16611: contig of 1100 bp in length
* 16612 17111: gap of unknown length
* 17112 17735: contig of 1024 bp in length
* 17736 17835: gap of unknown length
* 17836 19257: contig of 1422 bp in length
* 19258 19357: gap of unknown length
*
21141: contig of 1784 bp in length
21241: gap of unknown length
22719: contig of 1478 bp in length
22820 23882: contig of 1063 bp in length
23883 23982: gap of unknown length
23983 25414: contig of 1432 bp in length
25415 26921: contig of 1407 bp in length
26922 27021: gap of unknown length
27022 28161: contig of 1140 bp in length
28162 28261: gap of unknown length
28262 28663: contig of 1602 bp in length
28664 29964: gap of unknown length
29964 31473: contig of 1509 bp in length
31473 31572: gap of unknown length
31573 32709: contig of 1136 bp in length
32709 34029: gap of unknown length
34029 34129: contig of 1221 bp in length
34130 35944: gap of unknown length
35944 37311: contig of 1267 bp in length
37311 38831: contig of 1421 bp in length
38832 39311: gap of unknown length
39311 40451: contig of 1520 bp in length
40452 40551: gap of unknown length
40552 41719: contig of 1168 bp in length
41720 41819: gap of unknown length
41820 43453: contig of 1634 bp in length
43454 43553: gap of unknown length
43554 44779: contig of 1226 bp in length
44780 46754: contig of 1875 bp in length
46755 46854: gap of unknown length
46855 48580: contig of 1726 bp in length
48581 48680: gap of unknown length
48681 49932: contig of 1252 bp in length
49933 50032: gap of unknown length
50033 52279: contig of 2247 bp in length
52280 53279: gap of unknown length
53280 53623: contig of 1144 bp in length
53624 55547: contig of 1924 bp in length
55548 55647: gap of unknown length
55648 56883: contig of 1236 bp in length
56884 56983: gap of unknown length
56984 58628: contig of 1645 bp in length
58629 58728: gap of unknown length
58729 61144: contig of 2416 bp in length
61145 61244: gap of unknown length
61245 62701: contig of 1457 bp in length
62702 62801: gap of unknown length
62802 64563: contig of 1762 bp in length
64564 64663: gap of unknown length
64664 65604: contig of 1841 bp in length
65605 66505: gap of unknown length
66506 68567: contig of 1963 bp in length
68568 68667: gap of unknown length
68668 70346: contig of 1679 bp in length
70347 70446: gap of unknown length
70447 71987: contig of 1541 bp in length
71988 72087: gap of unknown length
72088 73462: contig of 1375 bp in length
73463 73562: gap of unknown length
73563 75187: contig of 1625 bp in length
75188 75287: gap of unknown length
75288 77198: contig of 1911 bp in length
77199 77298: gap of unknown length
77299 79285: contig of 1987 bp in length
79286 79386: gap of unknown length
81653: contig of 2268 bp in length

```

```

* 81654 81753: gap of unknown length
* 81754 83926: contig of 2173 bp in length
* 83927 84036: gap of unknown length
* 84037 86352: contig of 2326 bp in length
* 86353 86452: gap of unknown length
* 86453 89938: contig of 3486 bp in length
* 89939 90038: gap of unknown length

Query Match      0.6%; Score 21; DB 2; Length 150055;
Best Local Similarity 100.0%; Pred. NO. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2988 TCAGGAAGTGAAGGACAAAG 3008
      |||
Db 91054 TCAGGAAGTGAAGGACAAAG 91074

RESULT 27
AC128389
LOCUS
DEFINITION
Rattus norvegicus clone CH230-137B13, linear HTG 19-JUL-2002
*** 65 unordered pieces.
AC128389
VERSION
HTG: HTGS_PHASE1
KEYWORDS
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 153258)
Muzny,D.M., Adams,C., Adio-oduola,B., Ali-oshan,F.R., Allen,C.,
Alsbrooks,S.L., Amaraturge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbata,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokkwo,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Roife,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished

TITLE
JOURNAL

```

```

REFERENCE
2 (bases 1 to 153258)
Worley,K.C.
AUTHORS
Direct Submission
TITLE
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
JOURNAL
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GZAF
Center clone name: CH230-137B13
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 101481 bases at least Q40
Consensus quality: 109092 bases at least Q30
Consensus quality: 113403 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 65 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1154: contig of 1154 bp in length
1155: gap of unknown length
1255: contig of 1122 bp in length
2377: gap of unknown length
2476: gap of unknown length
3476: contig of 1000 bp in length
3477: gap of unknown length
3576: gap of unknown length
4585: contig of 1009 bp in length
4686: gap of unknown length
5732: contig of 1047 bp in length
5733: gap of unknown length
5833: contig of 1120 bp in length
5952: gap of unknown length
6953: contig of 1485 bp in length
7053: gap of unknown length
8538: contig of 1460 bp in length
10098: gap of unknown length
10197: gap of unknown length
11420: contig of 1223 bp in length
11421: gap of unknown length
11520: gap of unknown length
11521: contig of 1265 bp in length
12785: gap of unknown length
12886: gap of unknown length
14070: contig of 1185 bp in length
14071: gap of unknown length
14171: contig of 1133 bp in length
15303: gap of unknown length
15304: contig of 1693 bp in length
15404: gap of unknown length
17097: contig of 2026 bp in length
17197: gap of unknown length
19222: contig of 1479 bp in length
19223: gap of unknown length
19323: contig of 1415 bp in length
20802: gap of unknown length
20901: contig of 1288 bp in length
22317: gap of unknown length
22416: contig of 1288 bp in length
23645: gap of unknown length
23744: contig of 2040 bp in length
23745: gap of unknown length
25885: contig of 1607 bp in length
27491: gap of unknown length
27492: contig of 1318 bp in length
27592: gap of unknown length
28910: contig of 1722 bp in length
30731: gap of unknown length

```

30732 30831: gap of unknown length  
30832 contig of 1675 bp in length  
32507 32608: gap of unknown length  
32608 contig of 1421 bp in length  
34127 34127: gap of unknown length  
34128 35986: contig of 1859 bp in length  
35987 36086: gap of unknown length  
36087 37203: contig of 1117 bp in length  
37204 37303: gap of unknown length  
37304 38374: contig of 1071 bp in length  
38375 38474: gap of unknown length  
38475 40134: contig of 1660 bp in length  
40135 40234: gap of unknown length  
40235 41347: contig of 1113 bp in length  
41348 41447: gap of unknown length  
41448 43042: contig of 1595 bp in length  
43043 43142: gap of unknown length  
43143 44544: contig of 1402 bp in length  
44545 44644: gap of unknown length  
44645 46156: contig of 1512 bp in length  
46157 46256: gap of unknown length  
46257 47889: contig of 1633 bp in length  
47890 47989: gap of unknown length  
47990 50678: contig of 2689 bp in length  
50679 50778: gap of unknown length  
50779 52297: contig of 1519 bp in length  
52298 52397: gap of unknown length  
52398 53740: contig of 1343 bp in length  
53741 53840: gap of unknown length  
53841 55971: contig of 2131 bp in length  
55972 56071: gap of unknown length  
56073 56884: contig of 2613 bp in length  
56885 56984: gap of unknown length  
56985 60633: contig of 1849 bp in length  
60634 60733: gap of unknown length  
60734 62958: contig of 2225 bp in length  
62959 63058: gap of unknown length  
63059 65961: contig of 2903 bp in length  
65962 66061: gap of unknown length  
66062 68823: contig of 2762 bp in length  
68824 68923: gap of unknown length  
68924 70367: contig of 1444 bp in length  
70368 70467: gap of unknown length  
70468 72722: contig of 2235 bp in length  
72723 72822: gap of unknown length  
72823 75859: contig of 3037 bp in length  
75860 75959: gap of unknown length  
75960 78368: contig of 2409 bp in length  
78369 78468: gap of unknown length  
78469 80808: contig of 2340 bp in length  
80809 80908: gap of unknown length  
80909 83391: contig of 2483 bp in length  
83392 83491: gap of unknown length  
83492 85850: contig of 2359 bp in length  
85851 85951: gap of unknown length  
85951 88321: contig of 2371 bp in length  
88322 88421: gap of unknown length  
88422 91224: contig of 2703 bp in length  
91225 91224: gap of unknown length  
91225 93774: contig of 2550 bp in length  
93775 93874: gap of unknown length  
93875 96482: contig of 2608 bp in length  
96483 96582: gap of unknown length  
96583 97223: contig of 3141 bp in length  
97224 99823: gap of unknown length  
99824 103588: contig of 3765 bp in length  
103589 103688: gap of unknown length  
103689 106777: contig of 3089 bp in length  
106778 106877: gap of unknown length  
106878 110990: contig of 4113 bp in length  
110991 111090: gap of unknown length  
111091 114334: contig of 3244 bp in length  
114335 114434: gap of unknown length

Query Match 0.6%; Score 21; DB 2; Length 153258;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 997 TCCTACAGATTGTTCTCAA 1017  
Db 131524 TCCTACAGATTGTTCTCAA 131544  

RESULT 28  
AC121213 155859 bp DNA linear HTG 24-JUL-2002  
LOCUS Rattus norvegicus clone CH230-302H2, \*\*\* SEQUENCING IN PROGRESS  
DEFINITION \*\*\* 62 unordered pieces.  
AC121213  
AC121213.2 GI:21909448  
VERSION HTG: HTGS\_PHASE1.  
KEYWORDS Rattus norvegicus  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 155859)  
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhaq,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denna,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Hernandez,J., Harris,K., Hart,M., Havlak,P., Hayes,A., Hernandez,J., Homsif.F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,B., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Lieu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwkw,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N., Slason,I., Sodergren,E., Sonaake,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 155859)  
AUTHORS Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
REFERENCE 3 (bases 1 to 155859)



AUTHORS  
TITLE  
JOURNAL  
COMMENT

Worley, K.C.  
Direct Submission  
Submitted (24-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 19, 2002 this sequence version replaced gi:20806243.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GYLF  
Center clone name: CH230-302H2  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 105802 bases at least Q40  
Consensus quality: 112567 bases at least Q30  
Consensus quality: 116059 bases at least Q20  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 62 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
\* 1058: contig of 1058 bp in length  
\* 1059  
\* 1158: gap of unknown length  
\* 1159  
\* 2256: contig of 1098 bp in length  
\* 2257  
\* 2356: gap of unknown length  
\* 2357  
\* 3389: contig of 1033 bp in length  
\* 3390  
\* 3489: gap of unknown length  
\* 3490  
\* 4531: contig of 1042 bp in length  
\* 4532  
\* 4631: gap of unknown length  
\* 4632  
\* 5694: contig of 1063 bp in length  
\* 5695  
\* 5794: gap of unknown length  
\* 5795  
\* 6836: contig of 1042 bp in length  
\* 6837  
\* 6936: gap of unknown length  
\* 6937  
\* 8415: contig of 1479 bp in length  
\* 8416  
\* 8515: gap of unknown length  
\* 8516  
\* 9750: contig of 1235 bp in length  
\* 9751  
\* 9850: gap of unknown length  
\* 9851  
\* 11533: contig of 1683 bp in length  
\* 11534  
\* 11633: gap of unknown length  
\* 11634  
\* 12715: contig of 1082 bp in length  
\* 12716  
\* 12815: gap of unknown length  
\* 12816  
\* 14283: contig of 1468 bp in length  
\* 14284  
\* 14383: gap of unknown length  
\* 14384  
\* 15475: contig of 1092 bp in length  
\* 15476  
\* 15575: gap of unknown length  
\* 15576  
\* 16881: contig of 1306 bp in length  
\* 16882  
\* 16981: gap of unknown length  
\* 16982  
\* 18003: contig of 1022 bp in length  
\* 18004  
\* 18103: gap of unknown length  
\* 18104  
\* 19391: contig of 1288 bp in length  
\* 19392  
\* 19491: gap of unknown length  
\* 19492  
\* 20800: contig of 1309 bp in length  
\* 20801  
\* 20900: gap of unknown length  
\* 20901  
\* 22705: contig of 1805 bp in length  
\* 22706  
\* 22805: gap of unknown length  
\* 22806  
\* 23833: contig of 1028 bp in length  
\* 23834  
\* 23933: gap of unknown length  
\* 23934  
\* 25955: contig of 2022 bp in length  
\* 25956  
\* 26055: gap of unknown length  
\* 26056  
\* 28403: contig of 2348 bp in length  
\* 28404  
\* 28503: gap of unknown length  
\* 30037: contig of 1594 bp in length  
\* 28504

\* 30098  
\* 30198  
\* 32452: contig of 2255 bp in length  
\* 32453  
\* 32552: gap of unknown length  
\* 32553  
\* 33844: contig of 1292 bp in length  
\* 33845  
\* 33944: gap of unknown length  
\* 33945  
\* 36069: contig of 2125 bp in length  
\* 36070  
\* 36170  
\* 37965: contig of 1796 bp in length  
\* 37966  
\* 38065: gap of unknown length  
\* 38066  
\* 40084: contig of 2019 bp in length  
\* 40085  
\* 40184: gap of unknown length  
\* 41606: contig of 1422 bp in length  
\* 41607  
\* 41707  
\* 43596: contig of 1890 bp in length  
\* 43597  
\* 43696: gap of unknown length  
\* 43697  
\* 45243: contig of 1547 bp in length  
\* 45244  
\* 45343: gap of unknown length  
\* 45344  
\* 46995: contig of 1552 bp in length  
\* 46996  
\* 46995: gap of unknown length  
\* 46996  
\* 48736: contig of 1741 bp in length  
\* 48737  
\* 48836: gap of unknown length  
\* 48837  
\* 50741: contig of 1905 bp in length  
\* 50742  
\* 50841: gap of unknown length  
\* 50842  
\* 52954: contig of 2113 bp in length  
\* 52955  
\* 53054: gap of unknown length  
\* 53055  
\* 55087: contig of 2033 bp in length  
\* 55088  
\* 55187: gap of unknown length  
\* 55188  
\* 56968: contig of 1781 bp in length  
\* 56969  
\* 56968: gap of unknown length  
\* 57069  
\* 59180: contig of 2112 bp in length  
\* 59181  
\* 59280: gap of unknown length  
\* 59281  
\* 61244: contig of 1964 bp in length  
\* 61245  
\* 63444: gap of unknown length  
\* 63445  
\* 63657: contig of 2313 bp in length  
\* 63658  
\* 63757: gap of unknown length  
\* 63758  
\* 66400: contig of 2643 bp in length  
\* 66401  
\* 66500: gap of unknown length  
\* 66501  
\* 68431: contig of 1931 bp in length  
\* 68432  
\* 68531: gap of unknown length  
\* 68532  
\* 72267: contig of 3736 bp in length  
\* 72268  
\* 72367: gap of unknown length  
\* 72368  
\* 75276: contig of 2909 bp in length  
\* 75277  
\* 75376: gap of unknown length  
\* 75377  
\* 77301: contig of 1925 bp in length  
\* 77302  
\* 77401: gap of unknown length  
\* 77402  
\* 80313: contig of 2912 bp in length  
\* 80314  
\* 80413: gap of unknown length  
\* 80414  
\* 82990: contig of 2477 bp in length  
\* 82991  
\* 82990: gap of unknown length  
\* 82991  
\* 85535: contig of 2545 bp in length  
\* 85536  
\* 85635: gap of unknown length  
\* 85636  
\* 86017: contig of 2382 bp in length  
\* 86018  
\* 88117: gap of unknown length  
\* 88118  
\* 91242: contig of 3125 bp in length  
\* 91243  
\* 91342: gap of unknown length  
\* 91343  
\* 94958: contig of 3516 bp in length  
\* 94959  
\* 94958: gap of unknown length  
\* 94959  
\* 98762: contig of 3804 bp in length  
\* 98763  
\* 98862: gap of unknown length  
\* 98863  
\* 102013: contig of 3151 bp in length  
\* 102014  
\* 102113: gap of unknown length  
\* 102114  
\* 105285: contig of 3172 bp in length  
\* 105286  
\* 105385: gap of unknown length  
\* 105386  
\* 110193: contig of 4808 bp in length  
\* 110194  
\* 110293: gap of unknown length  
\* 110294  
\* 114524: contig of 4231 bp in length

Query Match 0.6%; Score 21; DB 2; Length 155859;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 581 TTTTTCCTTTTCTCTAGT 601  
|||||

Db 75430 TTTTTCCTTTTACTACT 75450

# RESULT 29

AC015953

LOCUS

DEFINITION

AC015953

VERSION

HTG: HTGS-PHASE1; HTGS-DRAFT.

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 155904)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Unpublished

2 (bases 1 to 155904)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,

Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,

Brown,A., Castle,A., Collangelo,M., Collins,S., Collamore,A.,

Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,

Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,

Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,

McEwan,P., McCurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,

Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Seaver,P.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,

Wyman,D., Ye-W.J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 28, 2000 this sequence version replaced gi:7382672.

All repeats were identified using RepeatMasker.

Smit, A.F.A. & Green, P. (1995-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: I953

Center clone name: 7\_M5

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-primer-amersham; 7% of reads

Chemistry: Dye-terminator Big Dye; 93% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 144110 bases at least Q40

Consensus quality: 147126 bases at least Q30

Consensus quality: 152138 bases at least Q20

Insert size: 168000; agarose-fp

Insert size: 153704; sum-of-contigs

Quality coverage: 3.7 in Q20 bases; agarose-fp

Quality coverage: 4.0 in Q20 bases

NOTE: This is a 'working draft' sequence. It currently

consists of 23 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 1321: contig of 1321 bp in length

1322 1421: gap of 100 bp

1422 2714: contig of 1293 bp in length

2715 2814: gap of 100 bp

-----

2815 4104: contig of 1290 bp in length

4105 4204: gap of 100 bp

4205 6487: contig of 2283 bp in length

6488 6587: gap of 100 bp

6588 8437: contig of 1750 bp in length

8438 11781: contig of 3344 bp in length

11782 11881: gap of 100 bp

11882 14350: contig of 2469 bp in length

14351 14450: gap of 100 bp

14451 19103: contig of 4652 bp in length

19103 19202: gap of 100 bp

19203 22602: contig of 3400 bp in length

22603 22702: gap of 100 bp

22703 26029: contig of 3327 bp in length

26030 26129: gap of 100 bp

26130 29282: contig of 3153 bp in length

29283 29382: gap of 100 bp

29383 34427: contig of 5045 bp in length

34428 34527: gap of 100 bp

34528 38012: contig of 3485 bp in length

38013 38112: gap of 100 bp

38113 43120: contig of 5008 bp in length

43121 43220: gap of 100 bp

43221 45302: contig of 6082 bp in length

45303 49402: gap of 100 bp

49403 55105: contig of 6703 bp in length

55106 56205: gap of 100 bp

56206 64386: contig of 8181 bp in length

64387 64486: gap of 100 bp

64487 71512: contig of 7026 bp in length

71513 71612: gap of 100 bp

71613 79328: contig of 7716 bp in length

79329 79428: gap of 100 bp

79429 96099: contig of 16671 bp in length

96100 96199: gap of 100 bp

96200 114101: contig of 17902 bp in length

114102 114201: gap of 100 bp

114202 134807: contig of 20606 bp in length

134808 134907: gap of 100 bp

134908 155904: contig of 20997 bp in length.

----- Location/Qualifiers

1. 155904

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="18"

/map="18"

/clone="RP11-7M5"

/clone\_lib="RPC1-11 Human Male BAC"

1. 1321

/note="assembly\_fragment"

1422. 2714

/note="assembly\_fragment"

2815. 4104

/note="assembly\_fragment"

4205. 6487

/note="assembly\_fragment"

6588. 8337

/note="assembly\_fragment"

clone\_end:SP6

vector\_side:right"

8438. 11781

/note="assembly\_fragment"

11882. 14350

/note="assembly\_fragment"

14451. 19102

/note="assembly\_fragment"

19203. 22602

/note="assembly\_fragment"

22703. 26029

/note="assembly\_fragment"

26130. 29282

/note="assembly\_fragment"

## FEATURES

source

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature 29383..34427  
/note="assembly\_fragment"  
34528..38012  
/note="assembly\_fragment"  
misc\_feature 38113..43120  
/note="assembly\_fragment"  
misc\_feature 43221..49302  
/note="assembly\_fragment"  
49403..56105  
/note="assembly\_fragment"  
misc\_feature 56206..64386  
/note="assembly\_fragment"  
misc\_feature 64487..71512  
/note="assembly\_fragment"  
misc\_feature 71613..79328  
/note="assembly\_fragment"  
79429..96099  
/note="assembly\_fragment"  
misc\_feature 96200..114101  
/note="assembly\_fragment"  
misc\_feature 114202..134807  
/note="assembly\_fragment"  
misc\_feature 134908..155904  
/note="assembly\_fragment"  
BASE COUNT 47807 a 29672 c 29375 g 46845 t 2205 others  
ORIGIN  
Query Match 0.6%: Score 21; DB 2; Length 155904;  
Best Local Similarity 100.0%: Pred. No. 27;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2329 GGAATATGCTCAGGTTTGT 2349  
|||||  
Db 63859 GGAATATGCTCAGGTTTGT 63879  
RESULT 30  
AC114071  
LOCUS  
DEFINITION Rattus norvegicus clone CH230-14017, \*\*\* SEQUENCING IN PROGRESS  
AC114071 157759 bp DNA linear HTG 13-JUL-2002  
VERSION AC114071.2 GI:21731941  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus  
1 (bases 1 to 157759)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbacia,J., Benton,J., Blincoe,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyie,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
Homsji,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlsson,E., Kelly,S., Khap,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseghe,H.,  
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,

10864 10963: gap of unknown length  
10964 12992: contig of 1329 bp in length  
12992 12992: gap of unknown length  
12993 13409: contig of 1017 bp in length  
13409 13509: gap of unknown length  
13509 15073: contig of 1564 bp in length  
15073 15173: gap of unknown length  
15173 15174: contig of 1416 bp in length  
15174 16589: gap of unknown length  
16589 16689: gap of unknown length  
16689 18332: contig of 1643 bp in length  
18332 18432: gap of unknown length  
18432 19956: contig of 1524 bp in length  
19956 20056: gap of unknown length  
20056 21151: contig of 1095 bp in length  
21151 21251: gap of unknown length  
21251 22251: contig of 1000 bp in length  
22251 22351: gap of unknown length  
22351 23738: contig of 1387 bp in length  
23738 23838: gap of unknown length  
23838 25427: contig of 1589 bp in length  
25427 25527: gap of unknown length  
25527 27387: contig of 1860 bp in length  
27387 27487: gap of unknown length  
27487 28922: contig of 1435 bp in length  
28922 29022: gap of unknown length  
29022 30162: contig of 1140 bp in length  
30162 31771: contig of 1509 bp in length  
31771 31871: gap of unknown length  
31871 34314: contig of 2443 bp in length  
34314 34414: gap of unknown length  
34414 35957: contig of 1543 bp in length  
35957 36057: gap of unknown length  
36057 38356: contig of 2299 bp in length  
38356 38456: gap of unknown length  
38456 39903: contig of 1447 bp in length  
39903 40003: gap of unknown length  
40003 41899: contig of 1696 bp in length  
41899 41799: gap of unknown length  
41799 43000: contig of 1201 bp in length  
43000 43100: gap of unknown length  
43100 44542: contig of 1442 bp in length  
44542 44642: gap of unknown length  
44642 46604: contig of 1962 bp in length  
46604 46704: gap of unknown length  
46704 48602: contig of 1898 bp in length  
48602 48702: gap of unknown length  
48702 51016: contig of 2314 bp in length  
51016 51116: gap of unknown length  
51116 53668: contig of 2552 bp in length  
53668 53768: gap of unknown length  
53768 55640: contig of 1872 bp in length  
55640 55740: gap of unknown length  
55740 58119: contig of 2379 bp in length  
58119 58219: gap of unknown length  
58219 59888: contig of 1669 bp in length  
59888 61976: contig of 1988 bp in length  
61976 62076: gap of unknown length  
62076 64583: contig of 2507 bp in length  
64583 64683: gap of unknown length  
64683 67559: contig of 2876 bp in length  
67559 67659: gap of unknown length  
67659 71558: contig of 3899 bp in length  
71558 71658: gap of unknown length  
71658 73750: contig of 2092 bp in length  
73750 73850: gap of unknown length  
73850 77404: contig of 3554 bp in length  
77404 77504: gap of unknown length  
77504 80262: contig of 2758 bp in length  
80262 80362: gap of unknown length  
80362 82789: contig of 2427 bp in length  
82789 82889: gap of unknown length

\* 82890 85869: contig of 2980 bp in length  
\* 85870 85969: gap of unknown length  
\* 85970 89739: contig of 3770 bp in length  
\* 89740 89839: gap of unknown length  
\* 89840 92739: contig of 2890 bp in length  
\* 92730 92839: gap of unknown length  
\* 92830 94946: contig of 2117 bp in length  
\* 94947 95046: gap of unknown length  
\* 95047 97684: contig of 2638 bp in length  
\* 97685 97784: gap of unknown length  
\* 97785 100890: contig of 3106 bp in length  
\* 100891 100990: gap of unknown length  
\* 100991 103985: contig of 2975 bp in length  
\* 103986 104065: gap of unknown length  
\* 104066 106355: contig of 2290 bp in length  
\* 106356 106455: gap of unknown length  
\* 106456 111448: contig of 4993 bp in length  
\* 111449 111548: gap of unknown length  
\* 111549 115697: contig of 4149 bp in length

Query Match 0.6%; Score 21; DB 2; Length 157759;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 581 TTTTTCCTTTTATCTAGT 601  
Db 137270 TTTTTCCTTTTATCTAGT 137290

RESULT 31

AC034304 AC034304 158158 bp DNA linear HTG 22-MAY-2000  
LOCUS AC034304  
DEFINITION Homo sapiens chromosome 8 clone RP11-194B7 map 8, WORKING DRAFT  
ACCESSION AC034304  
VERSION AC034304.2 GI:8018117  
KEYWORDS HTG; HTGS\_Phrase1; HTGS\_DRAFT.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 158158)  
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
TITLE Homo sapiens chromosome 8, clone RP11-194B7  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 158158)  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bada, F., Boguslavsky, L., Boukhgaiter, B., Brown, A., Burkett, G., Campopiano, A., Castler, A., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Milhova, T., Miranda, C., Mlenda, V., Morrow, J., Murphy, T., Navlor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierse, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schaefer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, C., Zainoun, J., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (05-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 22, 2000 this sequence version replaced gi.7417882.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)

```

http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Center: Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L9342
Center clone name: 194.B.7
-----
Summary Statistics
Sequencing vector: M13: M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 152650 bases at least Q40
Consensus quality: 155710 bases at least Q30
Consensus quality: 156663 bases at least Q20
Insert size: 160000; agarose-fp
Insert size: 137158; sum-of-contigs
Quality coverage: 4.9 in Q20 bases; agarose-fp
Quality coverage: 5.0 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1329: contig of 1329 bp in length
1330 1429: gap of 100 bp
1430 2703: contig of 1274 bp in length
2704 2803: gap of 100 bp
2804 4714: contig of 1911 bp in length
4715 4814: gap of 100 bp
4815 7418: contig of 2604 bp in length
7419 7518: gap of 100 bp
7519 10931: contig of 3413 bp in length
10932 11031: gap of 100 bp
11032 16459: contig of 5428 bp in length
16460 16599: gap of 100 bp
16560 29911: contig of 13352 bp in length
29912 30011: gap of 100 bp
30012 52449: contig of 22438 bp in length
52450 52549: gap of 100 bp
52550 83274: contig of 30725 bp in length
83275 83374: gap of 100 bp
83375 114699: contig of 31325 bp in length
114700 114799: gap of 100 bp
114800 158158: contig of 43359 bp in length.
Location/Qualifiers
1..158158
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-194B7"
/clone.lib="RPC1-11 Human Male BAC"
1..1339
/contig="assembly_fragment"
1430..2703
/contig="assembly_fragment"
2804..4714
/contig="assembly_fragment"
4815..7418
/contig="assembly_fragment"
7519..10931
/contig="assembly_fragment"
11032..16459
/contig="assembly_fragment"
16560..29911
/contig="assembly_fragment"
clone_end:T7

```

```

vector_side:right"
30012..52449
/contig="assembly_fragment"
52550..83274
/contig="assembly_fragment"
clone_end:SP6
vector_side:left"
83375..114699
/contig="assembly_fragment"
114800..158158
/contig="assembly_fragment"
BASE COUNT 49049 a 30381 c 29637 g 48091 t 1000 others
ORIGIN
Query Match 0.6%; Score 21; DB 2; Length 158158;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1941 AATCAGACACATAAAGTTAG 1961
|||||
Db 49858 AATCAGACACATAAAGTTAG 49878
|||||
RESULT 32
AC011155 158371 bp DNA linear PRI 17-FEB-2002
LOCUS Homo sapiens chromosome 18, clone RP11-7K9, complete sequence.
DEFINITION AC011155
ACCESSION AC011155
VERSION AC011155.13 GI:18698843
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
1 (bases 1 to 158371)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone RP11-7K9
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 158371)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeAtellano,K., Dewar,K., Domingo,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 158371)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeAtellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Lander,E., Lehoczky,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,

```

Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (16-OCR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 158371)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Canarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczyk, J., Levine, R., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

# TITLE JOURNAL

## COMMENT

Direct Submission  
Submitted (17-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Feb 17, 2002 this sequence version replaced gi:16152312.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIER

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L2947

Center clone name: 7\_K\_9

## FEATURES

### Source

Location/Qualifiers

1. .158371  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="18"  
/map="18"  
/clone="RP11-7K9"  
/clone\_lib="RP11 Human Male BAC"  
1184. .1231  
/rpt\_family="TTTGn"  
1607. .1627  
/rpt\_family="AT\_rich"  
2252. .2376  
/rpt\_family="MIR3"  
2974. .2996  
/rpt\_family="AT\_rich"  
complement(3290. .3497)  
/rpt\_family="MIR"  
complement(3639. .3676)  
/rpt\_family="MIR"  
3933. .3954  
/rpt\_family="AT\_rich"  
7421. .7462  
/rpt\_family="(CA)n"  
complement(8362. .8482)

### repeat\_region

### repeat\_region

### repeat\_region

### repeat\_region

### repeat\_region

### repeat\_region

### repeat\_region

### repeat\_region

### repeat\_region

### repeat\_region

### repeat\_region

### repeat\_region

### repeat\_region

### repeat\_region

### repeat\_region

/rpt\_family="L2"  
8734. .8780  
/rpt\_family="L2"  
complement(9127. .9470)  
/rpt\_family="THE1B"  
9827. .10183  
/rpt\_family="THE1B"  
complement(10775. .11105)  
/rpt\_family="MER1B"  
11147. .11568  
/rpt\_family="MSTA"  
12908. .13215  
/rpt\_family="ALUy"  
15012. .15611  
/rpt\_family="L2"  
15826. .15847  
/rpt\_family="AT\_rich"  
complement(19223. .19339)  
/rpt\_family="LMA2"  
19917. .19953  
/rpt\_family="(CA)n"  
21972. .21997  
/rpt\_family="AT\_rich"  
22094. .22189  
/rpt\_family="MIR"  
complement(22190. .22502)  
/rpt\_family="AluSx"  
22503. .22527  
/rpt\_family="MIR"  
complement(22546. .22677)  
/rpt\_family="MERSA"  
22622. .22692  
/rpt\_family="MERSA"  
23815. .23842  
/rpt\_family="(CA)n"  
24479. .24506  
/rpt\_family="AT\_rich"  
complement(24719. .24892)  
/rpt\_family="MER20"  
25709. .25753  
/rpt\_family="MIR"  
26575. .26610  
/rpt\_family="AT\_rich"  
27572. .27605  
/rpt\_family="polypyrimidine"  
27629. .27704  
/rpt\_family="L2"  
complement(27868. .28055)  
/rpt\_family="MERSA"  
28606. .28953  
/rpt\_family="L2"  
29786. .29986  
/rpt\_family="MIR"  
30391. .30421  
/rpt\_family="AT\_rich"  
31234. .31327  
/rpt\_family="(GAA)n"  
31684. .31744  
/rpt\_family="CT-rich"  
31849. .31990  
/rpt\_family="MIR3"  
32011. .32253  
/rpt\_family="MIR"  
complement(33243. .33390)  
/rpt\_family="MIR"  
34041. .34210  
/rpt\_family="L1PB3"  
34739. .34942  
/rpt\_family="MIR"  
complement(38566. .38717)  
/rpt\_family="MERSA"  
39098. .39251  
/rpt\_family="MIR"

```

repeat_region 39972..40277
/rpt_family="AluSg"
repeat_region 41183..43406
/rpt_family="L1P"
repeat_region 43582..43667
/rpt_family="L2"

```

Query Match 0.6% Score 21; DB 9; Length 158371;

Best Local Similarity 100.0%; Pred.No. 27; Mismatches 0; Indels 0; Gaps 0;

Qy 2329 GGAATATGTCAGGTTTGT 2349

Db 100964 GGAATATGTCAGGTTTGT 100984

RESULT 33

AC129367/c

LOCUS AC129367 160083 bp DNA linear HTG 02-AUG-2002  
 DEFINITION Rattus norvegicus clone CH230-253D15, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\* 53 unordered pieces.

ACCESSION AC129367

VERSION AC129367.1 GI:22001448

KEYWORDS HTG; HTGS-PHASE1.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 160083)

REFERENCE

AUTHORS

Munzy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Albrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
 Barbakid,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowie,S., Brivia,M., Brown,E., Brown,M., Bryant,N.P.,  
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Dearthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Fails,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
 Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
 Kratovic,S., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissege,H.,  
 Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
 Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwundu,G.,  
 Orgunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
 Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,  
 Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,  
 Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstock,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 160083)

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Worley,K.C.  
 Direct Submission  
 Submitted (29-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 160083)  
 Worley,K.C.  
 Direct Submission  
 Submitted (02-AUG-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GTOW  
 Center clone name: CH230-253D15  
 ----- Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye: 100% of reads  
 Assembly program: Phrap: version 0.990329  
 Consensus quality: 114920 bases at least Q40  
 Consensus quality: 120930 bases at least Q30  
 Consensus quality: 124914 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 53 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1072: contig of 1072 bp in length  
 \* 1073 1172: gap of unknown length  
 \* 1173 2556: contig of 1384 bp in length  
 \* 2557 2656: gap of unknown length  
 \* 2657 3986: contig of 1330 bp in length  
 \* 3987 4086: gap of unknown length  
 \* 4087 5599: contig of 1513 bp in length  
 \* 5600 5699: gap of unknown length  
 \* 5700 6744: contig of 1044 bp in length  
 \* 6744 8313: contig of unknown length  
 \* 8314 8413: gap of unknown length  
 \* 8414 9472: contig of 1059 bp in length  
 \* 9473 9572: gap of unknown length  
 \* 9573 10704: contig of 1131 bp in length  
 \* 10704 10803: gap of unknown length  
 \* 10804 12122: contig of 1319 bp in length  
 \* 12123 13406: contig of unknown length  
 \* 13407 13506: gap of unknown length  
 \* 13507 14978: contig of 1472 bp in length  
 \* 14979 15078: gap of unknown length  
 \* 15079 17119: contig of 2041 bp in length  
 \* 17120 17219: gap of unknown length  
 \* 17220 19195: contig of 1976 bp in length  
 \* 19196 19295: gap of unknown length  
 \* 19296 21926: contig of 2631 bp in length  
 \* 21927 23514: gap of unknown length  
 \* 23515 24835: contig of 1488 bp in length  
 \* 24836 24934: gap of unknown length  
 \* 24935 26513: contig of 1220 bp in length  
 \* 26514 27809: contig of 1579 bp in length  
 \* 27810 27909: contig of 1196 bp in length  
 \* 27910 27909: gap of unknown length

27910 29116: contig of 1207 bp in length  
29117 29216: gap of unknown length  
29217 31292: contig of 2076 bp in length  
31293 31392: gap of unknown length  
31394 32725: contig of 1333 bp in length  
32726 32825: gap of unknown length  
32826 35095: contig of 2270 bp in length  
35096 35195: gap of unknown length  
35196 38360: contig of 3165 bp in length  
38361 38460: gap of unknown length  
38461 40832: contig of 2372 bp in length  
40833 40932: gap of unknown length  
40933 43200: contig of 2268 bp in length  
43201 43300: gap of unknown length  
43301 45073: contig of 1773 bp in length  
45074 45173: gap of unknown length  
45174 48668: contig of 3495 bp in length  
48669 48768: gap of unknown length  
48769 51862: contig of 3094 bp in length  
51863 51962: gap of unknown length  
51963 53770: contig of 1808 bp in length  
53771 53870: gap of unknown length  
53871 55815: contig of 1945 bp in length  
55816 55915: gap of unknown length  
55916 58395: contig of 2480 bp in length  
58396 58495: gap of unknown length  
58496 61706: contig of 3211 bp in length  
61707 61806: gap of unknown length  
61807 64323: contig of 2517 bp in length  
64324 64423: gap of unknown length  
64424 67352: contig of 2829 bp in length  
67353 70001: contig of 2649 bp in length  
70002 70101: gap of unknown length  
70102 73603: contig of 3502 bp in length  
73604 73703: gap of unknown length  
73704 75880: contig of 2277 bp in length  
75881 81123: contig of 5043 bp in length  
81124 81224: gap of unknown length  
81224 86108: contig of 4885 bp in length  
86109 86208: gap of unknown length  
86209 89700: contig of 3492 bp in length  
89701 89800: gap of unknown length  
89801 92672: contig of 2872 bp in length  
92673 96858: contig of 4086 bp in length  
96859 96958: gap of unknown length  
96959 100450: contig of 3492 bp in length  
100451 100550: gap of unknown length  
100551 104087: contig of 3537 bp in length  
104088 104187: gap of unknown length  
104188 110817: contig of 6530 bp in length  
110818 110817: gap of unknown length  
110818 115054: contig of 4237 bp in length  
115055 115154: gap of unknown length  
115155 118921: contig of 3767 bp in length  
118922 119021: gap of unknown length  
119022 123715: contig of 4694 bp in length  
123716 123815: gap of unknown length  
123816 129153: contig of 5338 bp in length  
129154 129253: gap of unknown length  
129254 135387: contig of 6134 bp in length  
135388 135487: gap of unknown length  
135488 142433: contig of 6946 bp in length  
142434 142533: gap of unknown length  
142534 149464: contig of 6931 bp in length  
149465 149564: gap of unknown length  
149565 160083: contig of 10519 bp in length.

Location/Qualifiers  
1..160083  
/organism="Rattus norvegicus"

FEATURES  
Source

Query Match 0.6% Score 21; DB 2; Length 160083;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 581 TTTTTCCTTTTTCCTAGT 601  
|||||  
Db 60031 TTTTTCCTTTTTCCTAGT 60011  
  
RESULT 34  
AC091357 162600 bp DNA linear HTG 31-JUL-2002  
LOCUS Rattus norvegicus clone CH230-1A15, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
DEFINITION 77 unordered pieces.  
AC091357  
AC091357.4 GI:21953953  
VERSION HTG; HTGS\_PHASE1.  
KEYWORDS Norway rat.  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 162600)  
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbata,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Briefe,M., Brown,E., Brown,M., Bryant,N.P.,  
Buha,C., Burch,P., Burkett,C., Burrell,K.I., Byrd,N.C.,  
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
Homs,J.F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,  
Lozato,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Savery,G.,  
Rives,M., Rojas,A., Rojokan,I., Rolfe,M., Ruiz,S., Sisson,I.,  
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sison,I.,  
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Umani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
Williams,G., Williamson,A., Wleczek,R., Wooden,C., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,B.,  
Weinstock,G. and Gibbs,R.  
Direct Submission  
TITLE Unpublished  
JOURNAL 2 (bases 1 to 162600)  
AUTHORS Worley,K.C.  
REFERENCE Direct Submission  
TITLE Submitted (18-APR-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
JOURNAL 3 (bases 1 to 162600)  
AUTHORS Worley,K.C.



TITLE  
JOURNAL

Direct Submission  
Submitted (31-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 24, 2002 this sequence version replaced gi:21206167.

## COMMENT

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: TUEK  
Center clone name: CH230-LA15  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 55% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 82066 bases at least Q40  
Consensus quality: 88733 bases at least Q30  
Consensus quality: 91777 bases at least Q20  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 77 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1155: contig of 1155 bp in length  
\* 1156 1255: gap of unknown length  
\* 1256 2356: contig of 1101 bp in length  
\* 2357 2457: gap of unknown length  
\* 2458 3545: contig of 1089 bp in length  
\* 3546 3645: gap of unknown length  
\* 3646 4676: contig of 1031 bp in length  
\* 4677 5930: contig of 1154 bp in length  
\* 5931 6030: gap of unknown length  
\* 6031 7318: contig of 1288 bp in length  
\* 7319 7418: gap of unknown length  
\* 7419 8572: contig of 1154 bp in length  
\* 8573 8672: gap of unknown length  
\* 8673 9904: contig of 1231 bp in length  
\* 9904 10003: gap of unknown length  
\* 10003 11063: contig of 1060 bp in length  
\* 11064 11163: gap of unknown length  
\* 11164 12198: contig of 1035 bp in length  
\* 12199 12298: gap of unknown length  
\* 12299 13847: contig of 1549 bp in length  
\* 13848 13947: gap of unknown length  
\* 13948 14950: contig of 1003 bp in length  
\* 14951 15050: gap of unknown length  
\* 15051 16408: contig of 1358 bp in length  
\* 16409 16508: gap of unknown length  
\* 16509 17676: contig of 1168 bp in length  
\* 17677 17776: gap of unknown length  
\* 17777 18807: contig of 1031 bp in length  
\* 18808 18907: gap of unknown length  
\* 18908 20366: contig of 1459 bp in length  
\* 20367 20466: gap of unknown length  
\* 20467 21549: contig of 1083 bp in length  
\* 21550 21649: gap of unknown length  
\* 21650 22658: contig of 1009 bp in length  
\* 22659 22758: gap of unknown length  
\* 22759 24196: contig of 1437 bp in length  
\* 24196 24296: gap of unknown length  
\* 24296 25537: contig of 1242 bp in length  
\* 25538 25637: gap of unknown length

\* 25638 26918: contig of 1281 bp in length  
\* 26919 27018: gap of unknown length  
\* 27019 28091: contig of 1073 bp in length  
\* 28092 28191: gap of unknown length  
\* 28192 29567: contig of 1376 bp in length  
\* 29568 29667: gap of unknown length  
\* 29668 31186: contig of 1519 bp in length  
\* 31187 31286: gap of unknown length  
\* 31287 32779: contig of 1493 bp in length  
\* 32780 32879: gap of unknown length  
\* 32880 34510: contig of 1631 bp in length  
\* 34511 34610: gap of unknown length  
\* 34611 36269: contig of 1659 bp in length  
\* 36270 36369: gap of unknown length  
\* 36370 37749: contig of 1380 bp in length  
\* 37750 37849: gap of unknown length  
\* 37850 39396: contig of 1547 bp in length  
\* 39397 39496: gap of unknown length  
\* 39497 40701: contig of 1105 bp in length  
\* 40702 40801: gap of unknown length  
\* 40802 42561: contig of 1860 bp in length  
\* 42562 42661: gap of unknown length  
\* 42662 44434: contig of 1793 bp in length  
\* 44435 44534: gap of unknown length  
\* 44535 46043: contig of 1489 bp in length  
\* 46044 46143: gap of unknown length  
\* 46144 47705: contig of 1562 bp in length  
\* 47706 47805: gap of unknown length  
\* 47806 49580: contig of 1775 bp in length  
\* 49581 49680: gap of unknown length  
\* 49681 51450: contig of 1770 bp in length  
\* 51451 51550: gap of unknown length  
\* 51551 53143: contig of 1593 bp in length  
\* 53144 53243: gap of unknown length  
\* 53244 54806: contig of 1563 bp in length  
\* 54807 54906: gap of unknown length  
\* 54907 56515: contig of 1609 bp in length  
\* 56516 56615: gap of unknown length  
\* 56616 58202: contig of 1587 bp in length  
\* 58203 58303: gap of unknown length  
\* 58304 60165: contig of 1863 bp in length  
\* 60166 60265: gap of unknown length  
\* 60266 62601: contig of 2336 bp in length  
\* 62602 62701: gap of unknown length  
\* 62702 64499: contig of 1798 bp in length  
\* 64500 64599: gap of unknown length  
\* 64600 67052: contig of 2453 bp in length  
\* 67053 67152: gap of unknown length  
\* 67153 68917: contig of 1765 bp in length  
\* 68918 69017: gap of unknown length  
\* 69018 71403: contig of 2392 bp in length  
\* 71404 71503: gap of unknown length  
\* 71504 73977: contig of 2467 bp in length  
\* 73978 74076: gap of unknown length  
\* 74077 75585: contig of 1509 bp in length  
\* 75586 75685: gap of unknown length  
\* 75686 77579: contig of 1894 bp in length  
\* 77580 77679: gap of unknown length  
\* 77680 79357: contig of 1678 bp in length  
\* 79358 79457: gap of unknown length  
\* 79458 81312: contig of 1855 bp in length  
\* 81313 81412: gap of unknown length  
\* 81413 83721: contig of 2309 bp in length  
\* 83722 83821: gap of unknown length  
\* 83822 85452: contig of 1631 bp in length

Query Match 0.6%; Score 21; DB 2; Length 162600;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 997 TCCCTACAGATTGTTCAA 1017

DB 31768 TCCCTACAGATTGTTCAA 31788

```

RESULT 35
AP001802/c
LOCUS
DEFINITION
Homo sapiens chromosome 11 clone RP11-100J10 map 11q, WORKING DRAFT
SEQUENCE, 11 unordered pieces.
ACCESSION
AP001802
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
Homo sapiens DNA, clone:RP11-100J10.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE
Homo sapiens 165,699 genomic DNA of 11q
JOURNAL
Published Only in Database (2000)
REFERENCE
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (14-APR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@psc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)
On Aug 9, 2000 this sequence version replaced gi:8117473.
COMMENT
-----
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@psc.riken.go.jp
-----
Project Information
Center project name: HumDraff11
Center clone name: RP11-100J10
-----
Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 162162 bases at least Q40
Consensus quality: 163565 bases at least Q30
Consensus quality: 164180 bases at least Q20
Insert size: 164699; sum-of-contigs
Quality coverage: 11.49x in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of
11 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 39447 contig of 39447 bp in length
39548 39547: gap of 100 bp
71843 71742: contig of 32195 bp in length
99308 99207: contig of 27365 bp in length
118033 117932: contig of 18625 bp in length
13494 134393: contig of 16361 bp in length
13990 139889: contig of 5804 bp in length
13990 139889: contig of 5396 bp in length
145602 145501: contig of 5512 bp in length
150157 150056: contig of 100 bp
156061 156195: contig of 100 bp
162096 161995: contig of 100 bp
162096 165699: contig of 3604 bp in length.
FEATURES
source
1..165699
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="Rp11-100J10"
misc_feature 1..39447
/note="assembly_fragment"
misc_feature 39548..71742
/note="assembly_fragment"
misc_feature 71843..99207
/note="assembly_fragment"
misc_feature 99308..117932
/note="assembly_fragment"
misc_feature 118033..134393
/note="assembly_fragment"
misc_feature 13494..139889
/note="assembly_fragment clone_end:SP6 vector_side:right"
misc_feature 13990..145501
/note="assembly_fragment clone_end:T7 vector_side:left"
misc_feature 145602..150056
/note="assembly_fragment"
misc_feature 150157..155960
/note="assembly_fragment"
misc_feature 156061..161995
/note="assembly_fragment"
misc_feature 162096..165699
/note="assembly_fragment"
BASE COUNT 44435 a 35614 c 36787 g 47863 t 1000 others
ORIGIN
Query Match 0.6%; Score 21; DB 2; Length 165699;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 579 AATTTTTCCTTTTATCTA 599
|||||
Db 108062 AATTTTTCCTTTTATCTA 108042
|||||
RESULT 36
AC095109/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-7M2, *** SEQUENCING IN PROGRESS ***,
70 unordered pieces.
ACCESSION
AC095109
VERSION
AC095109.3 GI:21716907
KEYWORDS
HTG: HTGS_PHASE1.
SOURCE
Norway rat
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 168763)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbaria,J., Benton,J., Bimaga,K., Blarckenburg,K., Bonnin,D.,  
Buck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Bulay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Haves,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,  
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,  
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
Moser,N., Neal,D., Newton,J., Nwokwenkwo,S., Ogulu,M., Okwundu,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rives,M., Rojas,A., Rojibokan,I., Roife,M., Ruiz,S., Savery,G.,  
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,  
Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tatton,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.  
Direct Submission  
2 (bases 1 to 168763)  
Worley,K.C.  
Direct Submission  
Submitted (16-SEP-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 168763)  
Worley,K.C.  
Direct Submission  
Submitted (10-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 9, 2002 this sequence version replaced gi:17941975.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GCKQ  
Center clone name: CH230-7M2  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 107542 bases at least Q40  
Consensus quality: 115410 bases at least Q30  
Consensus quality: 122372 bases at least Q20

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 70 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1  
1194: contig of 1194 bp in length  
1294: gap of unknown length  
1295: contig of 1079 bp in length  
2373: gap of unknown length  
2374: contig of 1561 bp in length  
2474: gap of unknown length  
4035: contig of 1349 bp in length  
4135: gap of unknown length  
5483: contig of 1430 bp in length  
5584: gap of unknown length  
7013: contig of 1661 bp in length  
7113: gap of unknown length  
7114: contig of 1516 bp in length  
8775: gap of unknown length  
8775: contig of 1852 bp in length  
10391: gap of unknown length  
10391: contig of 1766 bp in length  
12343: gap of unknown length  
12442: contig of 1185 bp in length  
14209: gap of unknown length  
14309: contig of 1188 bp in length  
15494: gap of unknown length  
15594: gap of unknown length  
16781: gap of unknown length  
16881: contig of 1363 bp in length  
18244: gap of unknown length  
18245: contig of 1447 bp in length  
18345: gap of unknown length  
18345: contig of 1930 bp in length  
19792: gap of unknown length  
21821: contig of 1382 bp in length  
21921: gap of unknown length  
23303: contig of 1472 bp in length  
23403: gap of unknown length  
24875: gap of unknown length  
24876: contig of 1189 bp in length  
26164: gap of unknown length  
26165: contig of 2460 bp in length  
26255: gap of unknown length  
28724: contig of 1718 bp in length  
28725: gap of unknown length  
30542: contig of 1515 bp in length  
30642: gap of unknown length  
32157: gap of unknown length  
32157: contig of 2257 bp in length  
34514: gap of unknown length  
34614: contig of 1688 bp in length  
36302: gap of unknown length  
36402: contig of 1602 bp in length  
38004: gap of unknown length  
38104: contig of 1875 bp in length  
39979: gap of unknown length  
40079: gap of unknown length  
41784: contig of 1705 bp in length  
41884: gap of unknown length  
43604: contig of 1720 bp in length  
43704: gap of unknown length  
45685: contig of 1961 bp in length  
45685: gap of unknown length  
47591: contig of 1826 bp in length  
47691: gap of unknown length  
49059: contig of 1368 bp in length  
49159: gap of unknown length  
50579: contig of 1420 bp in length  
50679: gap of unknown length  
52301: contig of 1622 bp in length  
52401: gap of unknown length  
52302



109178 118993: contig of 9816 bp in length  
\* 118994 119093: gap of 100 bp  
\* 119094 128219: contig of 9126 bp in length  
\* 128220 128319: gap of 100 bp  
\* 128320 136675: contig of 8356 bp in length  
\* 136676 136775: gap of 100 bp  
\* 136776 144488: contig of 7713 bp in length  
\* 144489 144588: gap of 100 bp  
\* 144589 150077: contig of 5489 bp in length  
\* 150078 150177: gap of 100 bp  
\* 150178 156798: contig of 6621 bp in length  
\* 156799 156898: gap of 100 bp  
\* 156899 161857: contig of 4959 bp in length  
\* 161858 161957: gap of 100 bp  
\* 161958 165379: contig of 3422 bp in length  
\* 165380 165479: gap of 100 bp  
\* 165480 168313: contig of 2834 bp in length  
\* 168314 168413: gap of 100 bp  
\* 168414 169434: contig of 1021 bp in length  
\* 169435 169534: gap of 100 bp  
\* 169535 171196: contig of 1662 bp in length.  
FEATURES  
Source  
1. .171196  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="18"  
/map="18q12"  
/clone="RP11-683I7"  
misc\_feature  
1. .42291  
/note="assembly\_fragment"  
misc\_feature  
42392..63147  
/note="assembly\_fragment"  
misc\_feature  
63248..77851  
/note="assembly\_fragment"  
misc\_feature  
77952..88544  
/note="assembly\_fragment"  
misc\_feature  
88645..98724  
/note="assembly\_fragment"  
misc\_feature  
98825..109077  
/note="assembly\_fragment"  
misc\_feature  
109178..118993  
/note="assembly\_fragment"  
misc\_feature  
119094..128219  
/note="assembly\_fragment"  
misc\_feature  
128320..136675  
/note="assembly\_fragment"  
misc\_feature  
136776..144488  
/note="assembly\_fragment"  
misc\_feature  
144589..150077  
/note="assembly\_fragment"  
misc\_feature  
150178..156798  
/note="assembly\_fragment"  
misc\_feature  
156899..161857  
/note="assembly\_fragment"  
misc\_feature  
161958..165379  
/note="assembly\_fragment"  
misc\_feature  
165480..168313  
/note="assembly\_fragment"  
misc\_feature  
168414..169434  
/note="assembly\_fragment"  
misc\_feature  
169535..171196  
/note="assembly\_fragment"  
BASE COUNT 48636 a 35335 c 36221 g 49404 t 1600 others  
ORIGIN  
Query Match 0.6%; Score 21; DB 2; Length 171196;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 579 AATTTTCTTTTATCTA 599  
|||||  
Db 146855 AATTTTCTTTTATCTA 146835  
|||||

RESULT 38  
AP000880  
LOCUS  
DEFINITION  
Homo sapiens genomic DNA, chromosome 11q, clone:RP11-839D17,  
complete sequence.  
ACCESSION  
AP000880  
VERSION  
AP000880.4 GI:13359381  
KEYWORDS  
HTG.  
SOURCE  
Homo sapiens DNA, clone:RP11-839D17.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (sites)  
AUTHORS  
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE  
Homo sapiens genomic DNA  
JOURNAL  
Published Only in Database (1999)  
REFERENCE  
2 (bases 1 to 171853)  
AUTHORS  
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (13-DEC-1999) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:hattori@sc.riken.go.jp, URL:http://hgp.riken.go.jp/  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
COMMENT  
On Mar 16, 2001 this sequence version replaced gi:10130050.  
FEATURES  
Location/Qualifiers  
1..171853  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="11q"  
/clone="RP11-839D17"  
BASE COUNT 47378 a 38606 c 37154 g 48715 t  
ORIGIN  
Query Match 0.6%; Score 21; DB 9; Length 171853;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 579 AATTTTCTTTTATCTA 599  
|||||  
Db 51133 AATTTTCTTTTATCTA 51153  
|||||  
RESULT 39  
AC126420/c  
LOCUS  
DEFINITION  
Mus musculus clone RP23-153M4, WORKING DRAFT SEQUENCE, 30 ordered  
pieces.  
ACCESSION  
AC126420  
VERSION  
AC126420.1 GI:21699591  
KEYWORDS  
HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 175245)  
AUTHORS  
Birren,B., Nusbaum,C. and Lander,E.  
TITLE  
Mus musculus, clone RP23-153M4  
JOURNAL  
Unpublished  
REFERENCE  
2 (bases 1 to 175245)  
AUTHORS  
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
Barna.N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,  
Camaratia,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A., S.,  
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gord,S., Graham,L., Grand-pierre,N., Hagos,B.,  
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,

McCarthy M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,  
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,  
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
 Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,  
 Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,  
 Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,  
 Zembek, L., Zimmer, A. and Zody, M.

# TITLE JOURNAL

## COMMENT

Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: 153\_M\_4

Center clone name: 153\_M\_4

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 16436 bases at least Q40

Consensus quality: 169284 bases at least Q30

Consensus quality: 171255 bases at least Q20

Insert size: 183000; agarose-fp

Quality coverage: 4.8 in Q20 bases; agarose-fp

Quality coverage: 5.1 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently

consists of 30 contigs. Gaps between the contigs

are represented as runs of N. The order of the pieces

is believed to be correct as given, however the sizes

of the gaps between them are based on estimates that have

been provided by the submitter.

This sequence will be replaced

by the finished sequence as soon as it is available and

the accession number will be preserved.

1 2956: contig of 2956 bp in length

2957 3056: gap of 100 bp

3057 4129: contig of 1073 bp in length

4130 4229: gap of 100 bp

4230 5714: contig of 1485 bp in length

5715 5814: gap of 100 bp

5815 7318: contig of 1504 bp in length

7319 7418: gap of 100 bp

7419 8623: contig of 1205 bp in length

8624 8723: gap of 100 bp

8724 10175: contig of 1452 bp in length

10176 10275: gap of 100 bp

10276 11370: contig of 1095 bp in length

11371 11470: gap of 100 bp

11471 12614: contig of 1144 bp in length

12615 12714: gap of 100 bp

12715 14823: contig of 2109 bp in length

14824 14923: gap of 100 bp

14924 16075: contig of 1152 bp in length

16076 16175: gap of 100 bp

16176 48108: contig of 3193 bp in length

48109 48208: gap of 100 bp

48209 50534: contig of 2326 bp in length

50535 50634: gap of 100 bp

50635 52640: contig of 2006 bp in length

52641 52740: gap of 100 bp

52741 55523: contig of 2783 bp in length

55524 55623: gap of 100 bp

55624 58689: contig of 3066 bp in length

58690 61160: contig of 2371 bp in length

61161 61260: gap of 100 bp

61261 6311: contig of 2051 bp in length

6312 6341: gap of 100 bp

6342 65922: contig of 2511 bp in length

65923 66022: gap of 100 bp

66023 68151: contig of 2129 bp in length

68152 68251: gap of 100 bp

68252 70510: contig of 2259 bp in length

70511 70610: gap of 100 bp

70611 72725: contig of 2115 bp in length

72726 72825: gap of 100 bp

72826 74380: contig of 1555 bp in length

74381 74480: gap of 100 bp

74481 77461: contig of 2981 bp in length

77462 77561: gap of 100 bp

77562 89014: contig of 11453 bp in length

89015 89114: gap of 100 bp

89115 100866: contig of 11772 bp in length

100867 100986: gap of 100 bp

100987 116990: contig of 16004 bp in length

116991 117090: gap of 100 bp

117091 135503: contig of 18413 bp in length

135504 135603: gap of 100 bp

135604 154923: contig of 19320 bp in length

154924 155023: gap of 100 bp

155024 174663: contig of 19640 bp in length

174664 174763: gap of 100 bp

174764 175245: contig of 482 bp in length.

58690 58789: gap of 100 bp

58790 61160: contig of 2371 bp in length

61161 61260: gap of 100 bp

61261 6311: contig of 2051 bp in length

6312 6341: gap of 100 bp

6342 65922: contig of 2511 bp in length

65923 66022: gap of 100 bp

66023 68151: contig of 2129 bp in length

68152 68251: gap of 100 bp

68252 70510: contig of 2259 bp in length

70511 70610: gap of 100 bp

70611 72725: contig of 2115 bp in length

72726 72825: gap of 100 bp

72826 74380: contig of 1555 bp in length

74381 74480: gap of 100 bp

74481 77461: contig of 2981 bp in length

77462 77561: gap of 100 bp

77562 89014: contig of 11453 bp in length

89015 89114: gap of 100 bp

89115 100866: contig of 11772 bp in length

100867 100986: gap of 100 bp

100987 116990: contig of 16004 bp in length

116991 117090: gap of 100 bp

117091 135503: contig of 18413 bp in length

135504 135603: gap of 100 bp

135604 154923: contig of 19320 bp in length

154924 155023: gap of 100 bp

155024 174663: contig of 19640 bp in length

174664 174763: gap of 100 bp

174764 175245: contig of 482 bp in length.

## FEATURES

### Source

1. .175245  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="RP23-153M4"  
 /clone\_lib="RPCI-23 Female Mouse BAC"

### misc\_feature

1. .2956  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:left

### misc\_feature

3057..4129  
 /note="assembly\_fragment"

### misc\_feature

4230..5714  
 /note="assembly\_fragment"

### misc\_feature

5815..7318  
 /note="assembly\_fragment"

### misc\_feature

7419..8623  
 /note="assembly\_fragment"

### misc\_feature

8724..10175  
 /note="assembly\_fragment"

### misc\_feature

10276..11370  
 /note="assembly\_fragment"

### misc\_feature

11471..12614  
 /note="assembly\_fragment"

### misc\_feature

12715..14823  
 /note="assembly\_fragment"

### misc\_feature

14924..16075  
 /note="assembly\_fragment"

### misc\_feature

16176..48108  
 /note="assembly\_fragment"

### misc\_feature

48209..50534  
 /note="assembly\_fragment"

### misc\_feature

50635..52640  
 /note="assembly\_fragment"

### misc\_feature

52741..55523  
 /note="assembly\_fragment"

### misc\_feature

55624..58689  
 /note="assembly\_fragment"

### misc\_feature

58790..61160  
 /note="assembly\_fragment"

### misc\_feature

61261..63311  
 /note="assembly\_fragment"

### misc\_feature

63412..65922

```

misc_feature      /note="assembly_fragment"
66023..68151
/note="assembly_fragment"
68252..70510
/note="assembly_fragment"
70611..72725
/note="assembly_fragment"
72826..74380
/note="assembly_fragment"
74481..77451
/note="assembly_fragment"
77562..89014
/note="assembly_fragment"
89115..100886
/note="assembly_fragment"
100987..116990
/note="assembly_fragment"
117091..135503
/note="assembly_fragment"
135604..154923
/note="assembly_fragment"
155024..174663
/note="assembly_fragment"
174764..175245
/note="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT      51998 a 36415 c 35671 g 48251 t 2910 others
ORIGIN
Query Match      0.6%; Score 21; DB 2; Length 175245;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2915 AAAGAATACCCAGACTTCATG 2935
|||||
Db 81991 AAAGAATACCCAGACTTCATG 81971

RESULT 40
AP002840
LOCUS
DEFINITION      Homo sapiens genomic DNA, chromosome 11q, clone:RP11-159N11,
complete sequences.
ACCESSION      AP002840.2 GI:13488763
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Hattori.M., Ishii.K., Toyoda.A., Taylor.T.D., Hong-Seog.P.,
Fujiyama.A., Yada.T., Totoki.Y., Watanabe.H. and Sakaki.Y.
1 (sites)
Published Only in DataBase (2000)
Hattori.M., Ishii.K., Toyoda.A., Taylor.T.D., Hong-Seog.P.,
Fujiyama.A., Yada.T., Totoki.Y., Watanabe.H. and Sakaki.Y.
Direct Submission
Submitted (04-AUG-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
On Mar 29, 2001 this sequence version replaced gi:9757543.
COMMENT
FEATURES
Location/Qualifiers
1..177034
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-159N11"

```

```

BASE COUNT      48001 a 42063 c 41121 g 45849 t
ORIGIN
Query Match      0.6%; Score 21; DB 9; Length 177034;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1966 CTTGGATGGAGCAAAATATCA 1986
|||||
Db 81883 CTTGGATGGAGCAAAATATCA 81903

RESULT 41
AC009420
LOCUS
DEFINITION      Homo sapiens clone RP11-44J17, WORKING DRAFT SEQUENCE, 7 unordered
pieces.
ACCESSION      AC009420
VERSION
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 178229)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-44J17
Unpublished
2 (bases 1 to 178229)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferrelra,P., FitzHugh,W., Forrest,C.,
Funker,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris.W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (22-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 20, 2001 this sequence version replaced gi:7407920.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: Li079
Center clone name: 44_J-17
----- Summary Statistics
Sequencing vector: M13; M77815; 31% of reads
Sequencing vector: Plasmid; n/a; 69% of reads
Chemistry: Dye-primer-amersham; 2% of reads
Chemistry: Dye-terminator Big Dye; 98% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 175652 bases at least Q40
Consensus quality: 176863 bases at least Q30
Consensus quality: 177333 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 177629; sum-of-contigs
Quality coverage: 11.3 in Q20 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces

```





\*\*\*\*\*  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 14 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submittor.

\* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

\* 1 11403: contig of 11403 bp in length  
 \* 11404 11503: gap of 100 bp  
 \* 11504 12952: contig of 1449 bp in length  
 \* 12953 13052: gap of 100 bp  
 \* 13053 14155: contig of 1103 bp in length  
 \* 14156 14255: gap of 100 bp  
 \* 14256 16221: contig of 1966 bp in length  
 \* 16222 16321: gap of 100 bp  
 \* 16322 18465: contig of 2144 bp in length  
 \* 18466 18565: gap of 100 bp  
 \* 18566 22007: contig of 3442 bp in length  
 \* 22008 22107: gap of 100 bp  
 \* 22108 29764: contig of 7657 bp in length  
 \* 29765 29864: gap of 100 bp  
 \* 29865 38510: contig of 8646 bp in length  
 \* 38511 39610: gap of 100 bp  
 \* 39611 45758: contig of 7148 bp in length  
 \* 45759 45855: gap of 100 bp  
 \* 45859 60388: contig of 14530 bp in length  
 \* 60389 60488: gap of 100 bp  
 \* 60489 130153: contig of 69665 bp in length  
 \* 130154 130253: gap of 100 bp  
 \* 130254 148489: contig of 18236 bp in length  
 \* 148490 148589: gap of 100 bp  
 \* 148590 168118: contig of 19529 bp in length  
 \* 168119 168218: gap of 100 bp  
 \* 168219 179025: contig of 10807 bp in length.

## FEATURES

source

1. .179025  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="RP23-382N7"  
 /clone\_lib="RPC1-23 Female Mouse BAC"  
 1. .11403  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:left  
 11504..12952  
 /note="assembly\_fragment"  
 13053..14155  
 /note="assembly\_fragment"  
 14256..16221  
 /note="assembly\_fragment"  
 16322..18465  
 /note="assembly\_fragment"  
 18566..22007  
 /note="assembly\_fragment"  
 22108..29764  
 /note="assembly\_fragment"  
 29865..38510  
 /note="assembly\_fragment"  
 38611..45758  
 /note="assembly\_fragment"  
 45859..60388  
 /note="assembly\_fragment"  
 60489..130153  
 /note="assembly\_fragment"  
 130254..148489  
 /note="assembly\_fragment"  
 148590..168118  
 /note="assembly\_fragment"  
 168219..179025  
 /note="assembly\_fragment"

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

clone\_end:T7

vector\_side:right"

BASE COUNT 56670 a 32505 c 32335 g 56209 t 1306 others  
 ORIGIN

Query Match 0.6%; Score 21; DB 2; Length 179025;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 579 AATTTTTCCTTTTATCTA 599

Db 63553 AATTTTTCCTTTTATCTA 63533

RESULT 43

AP001934

LOCUS

DEFINITION

AP001934

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On May 30, 2000 this sequence version replaced gi:7678796.

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: http://hgp.gsc.riken.go.jp/

Contact: hattori@gsc.riken.go.jp

----- Project Information

Center project name: HumDraft18

Center clone name: RP11-711G5

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 158823 bases at least Q40

Consensus quality: 168306 bases at least Q30

Consensus quality: 172989 bases at least Q20

Insert size: 175581; sum-of-contigs

Quality coverage: 4.49x in Q20 bases; sum-of-contigs

-----

NOTE: This is a 'working draft' sequence. It currently consists of

38 contigs. The true order of the pieces is not known and their

order in this sequence record is arbitrary. Gaps between the

contigs are represented as runs N, but the exact sizes of the gaps

are unknown. This record will be updated with the finished sequence

as soon as it is available and the accession number will be

preserved

1 31182 contig of 31182 bp in length

31283 47814 contig of 16532 bp in length

47915 63599 contig of 15685 bp in length

63700 78592 contig of 14893 bp in length

78693 92264 contig of 13572 bp in length

92365 98324 contig of 5960 bp in length

AP001934 179281 bp DNA linear HTG 30-MAY-2000  
 Homo sapiens chromosome 18 clone RP11-711G5 map 18q21, WORKING  
 DRAFT SEQUENCE, 38 unordered pieces.

AP001934

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

Homo sapiens DNA, clone:RP11-711G5.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 179281)

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

Homo sapiens 179,281 genomic DNA of 18q21

Published only in DataBase (2000)

2 (bases 1 to 179281)

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

Direct Submission

Submitted (28-APR-2000) Masahira Hattori, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,

Japan (E-mail:hattori@gsc.riken.go.jp, Tel:81-42-778-9923,

URL:http://hgp.gsc.riken.go.jp/)

On May 30, 2000 this sequence version replaced gi:7678796.

----- Genome Center

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: http://hgp.gsc.riken.go.jp/

Contact: hattori@gsc.riken.go.jp

----- Project Information

Center project name: HumDraft18

Center clone name: RP11-711G5

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 158823 bases at least Q40

Consensus quality: 168306 bases at least Q30

Consensus quality: 172989 bases at least Q20

Insert size: 175581; sum-of-contigs

Quality coverage: 4.49x in Q20 bases; sum-of-contigs

-----

NOTE: This is a 'working draft' sequence. It currently consists of

38 contigs. The true order of the pieces is not known and their

order in this sequence record is arbitrary. Gaps between the

contigs are represented as runs N, but the exact sizes of the gaps

are unknown. This record will be updated with the finished sequence

as soon as it is available and the accession number will be

preserved

1 31182 contig of 31182 bp in length

31283 47814 contig of 16532 bp in length

47915 63599 contig of 15685 bp in length

63700 78592 contig of 14893 bp in length

78693 92264 contig of 13572 bp in length

92365 98324 contig of 5960 bp in length

[illegible]

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2329 GGAATATGTCAGGTTTGT 2349  
|||||  
Db 94307 GGAATATGTCAGGTTTGT 94327

RESULT 44  
AC083994  
LOCUS  
DEFINITION  
AC083994  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L11472  
Center clone name: 222\_H-22  
----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 175722 bases at least Q40  
Consensus quality: 178732 bases at least Q30  
Consensus quality: 179651 bases at least Q20  
Insert size: 189000; agarose-ff  
Insert size: 180065; sum-of-contigs  
Quality coverage: 5.4 in Q20 bases; agarose-ff  
Quality coverage: 5.7 in Q20 bases; sum-of-contigs  
-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces

\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 2629: contig of 2629 bp in length  
2630 2729: gap of 100 bp  
2730 2903: contig of 174 bp in length  
2904 3003: gap of 100 bp  
3004 4169: contig of 1166 bp in length  
4170 4269: gap of 100 bp  
4270 9015: contig of 4746 bp in length  
9016 9115: gap of 100 bp  
9116 16346: contig of 7231 bp in length  
16347 16446: gap of 100 bp  
16447 24593: contig of 8153 bp in length  
24600 24699: gap of 100 bp  
24700 34073: contig of 9374 bp in length  
34074 34173: gap of 100 bp  
34174 48313: contig of 14140 bp in length  
48314 48413: gap of 100 bp  
48414 74955: contig of 26542 bp in length  
74956 75055: gap of 100 bp  
75056 103827: contig of 28772 bp in length  
103828 103927: gap of 100 bp  
103928 133045: contig of 29118 bp in length  
133046 133145: gap of 100 bp  
133146 157551: contig of 24406 bp in length  
157552 157651: gap of 100 bp  
157652 181265: contig of 23614 bp in length.

FEATURES  
source  
1..181265  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-222H22"  
/clone\_lib="RPC1-11 Human Male BAC"  
1..2629  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left  
2730..2903  
/note="assembly\_fragment"  
3004..4169  
/note="assembly\_fragment"  
4270..9015  
/note="assembly\_fragment"  
9116..16346  
/note="assembly\_fragment"  
16447..24599  
/note="assembly\_fragment"  
24700..34073  
/note="assembly\_fragment"  
34174..48313  
/note="assembly\_fragment"  
48414..74955  
/note="assembly\_fragment"  
75056..103827  
/note="assembly\_fragment"  
103928..133045  
/note="assembly\_fragment"  
133146..157551  
/note="assembly\_fragment"  
157652..181265  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right

BASE COUNT 56264 a 34004 c 34091 g 55706 t 1200 others  
ORIGIN  
Query Match 0.6%; Score 21; DB 2; Length 181265;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1941 AATCAGACACATAAAGTTAG 1961  
|||||  
Db 117759 AATCAGACACATAAAGTTAG 117779

RESULT 45  
AC113269  
LOCUS  
DEFINITION Mus musculus clone RP23-372F2, WORKING DRAFT SEQUENCE, 13 ordered pieces  
AC113269  
VERSION HTG: HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 183951)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
JOURNAL Unpublished  
TITLE Submitted  
REFERENCE 2 (bases 1 to 183951)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,I., Colangelo,M., Collins,S., Collamore,A., Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severly,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
TITLE Direct Submission  
JOURNAL Submitted (28-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE 3 (bases 1 to 183951)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,I., Colangelo,M., Collins,S., Collamore,A., Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerard,M., Fitchugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Marquis,N., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severly,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
TITLE Direct Submission  
JOURNAL Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 6, 2002 this sequence version replaced gi:21313780.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L20332  
Center clone name: 372\_F-2  
----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 180116 bases at least Q40  
Consensus quality: 181644 bases at least Q30  
Consensus quality: 182175 bases at least Q20  
Insert size: 182000; agarose-fp  
Insert size: 182751; sum-of-contigs  
Quality coverage: 7.9 in Q20 bases; agarose-fp  
Quality coverage: 7.8 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 13 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\*  
\* 1 8760: contig of 8760 bp in length  
\* 8761 8860: gap of 100 bp  
\* 8861 10331: contig of 1471 bp in length  
\* 10332 10431: gap of 100 bp  
\* 10432 11525: contig of 1094 bp in length  
\* 11526 11625: gap of 100 bp  
\* 11626 13553: contig of 1928 bp in length  
\* 13554 13653: gap of 100 bp  
\* 13654 20393: contig of 6740 bp in length  
\* 20394 20493: gap of 100 bp  
\* 20494 25204: contig of 4711 bp in length  
\* 25205 25304: gap of 100 bp  
\* 25305 35649: contig of 10345 bp in length  
\* 35650 35749: gap of 100 bp  
\* 35750 55284: contig of 19535 bp in length  
\* 55285 55384: gap of 100 bp  
\* 55385 74290: contig of 18906 bp in length  
\* 74291 74390: gap of 100 bp  
\* 74391 98334: contig of 23944 bp in length  
\* 98335 98434: gap of 100 bp  
\* 98435 127342: contig of 28908 bp in length  
\* 127343 127443: gap of 100 bp  
\* 127443 148645: contig of 21203 bp in length  
\* 148646 148745: gap of 100 bp  
\* 148746 183951: contig of 35206 bp in length.  
Location/Qualifiers  
1. 183951  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-372F2"  
/clone\_lib="RPCI-23 Female Mouse BAC"  
1. 8760  
/note="assembly\_fragment"  
vector\_side:left  
clone\_end:sp6  
8861..10331  
/note="assembly\_fragment"  
10432..11525  
/note="assembly\_fragment"

FEATURES

source  
misc\_feature  
misc\_feature  
misc\_feature



```
-----
2157      2187      4631      4551      3608      3588
-----
6879      6905      117      <800      1567      1554
-----
15114     15028     3117     3078     3741     3711
-----
2779      2817      2549     2585     1952     1933
-----
65         <800      182      <800     3214     3186
-----
1945      1860      5443     5372     17299    17528
-----
4326      4363      8852     9062     3122     3186
-----
4652      4628      4125     4204     362      <800
-----
187        <800      933      923      4432     4389
-----
316        <800      1529     1540     2729     2795
-----
14460     14089     6315     6353     1751     1744
-----
2663      2694      1564     1540     1324     1304
-----
2093      2080      4331     4204     3915     3919
-----
9930      9994      3370     3419     3954     3919
-----
4370      4363      1951     2013     1572     1554
-----
513        <800      3853     3777     8737     8662
-----
2015      1994      9178     9062     8423     8376
-----
1536      1513      3482     3419     1959     1933
-----
4928      4930      6091     6353     36      <800
-----
220        <800      557      <800     1853     1848
-----
683        <800      3254     3419     1219     1201
-----
2804      2817      2546     2585     1474     1455
-----
881        889      6405     6353     16039    15844
-----
9220      9368      45      <800     2397     2405
-----
428        <800      807      812     2145     2138
-----
7185      7290      894      812     13246    13238
-----
2562      2560      4325     4204     1054     1089
-----
11284     11207     3049     3078     438     <800
-----
9486      9368      780      812     399     <800
-----
719        <800      40      <800     2662     2641
-----
1851      1860      2524     2585     7352     7426
-----
-----
3459      3419     6142     6153
-----
354        <800     926     931
-----
76         <800      214     <800
-----
1585      1540     2308     2302
-----
1150      1149     7936     7982
-----
2984      3078     695     <800
-----

Query Match      0.6%; Score 21; DB 9; Length 185148;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3395 TTGAAAGAGCTCATTTCATT 3415
|||||
Db 175598 TTGAAAGAGCTCATTTCATT 175618

RESULT 47
AC084367
LOCUS
DEFINITION Homo sapiens chromosome 2 clone RP11-359B5, WORKING DRAFT SEQUENCE,
33 unordered pieces.
ACCESSION AC084367
VERSION AC084367.2 GI:11039565
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS
REFERENCE 2 (bases 1 to 186386)
TITLE Direct Submission
AUTHORS Waterston,R.H.
JOURNAL Submitted (26-OCT-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Oct 30, 2000 this sequence version replaced gi:11024936.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0359B05
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-primer ET; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 166829 bases at least Q40
Consensus quality: 173228 bases at least Q30
Consensus quality: 176024 bases at least Q20
Insert size: 204000; agarose-fp
Insert size: 182402; sum-of-contigs
Quality coverage: 3.47 in Q20 bases; agarose-fp
Quality coverage: 3.77 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
```



```

ORIGIN
Query Match          0.6%, Score 21; DB 2; Length 186386;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 579 AATTTTTCCTTTTCTCTA 599
|||||
Db 86483 AATTTTTCCTTTTCTCTA 86503

RESULT 48
AC125777/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-7B13, *** SEQUENCING IN PROGRESS ***
AC125777
AC125777
AC125777.1 GI:21630402
VERSION
HTG; HTGS_PHASE1.
KEYWORDS
Norway rat.
SOURCE
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 189748)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Avele,M., Banks,T.,
Barbaria,J., Benton,J., Bimaga,K., Blackburn,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,S., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Doutwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgato,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,X., Johnson,K., Jolivet,S., Joudan,S.,
Karlisson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
L.J., Li,Z., Lichtarge,C., Lieu,C., Liu,J., Liu,W., Loulsegad,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Maxiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,S.,
Nguyen,N., Nickerson,B., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
Oraguine,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Woodson,S., Worley,K.,
Wu,C., Wu,Y., Wu,F.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 189748)
Worley,K.C.
Direct Submission
Submitted (30-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
3 (bases 1 to 189748)
Worley,K.C.
Direct Submission
Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GGCJ
Center clone name: CH230-7B13
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 125061 bases at least Q40
Consensus quality: 130947 bases at least Q30
Consensus quality: 136357 bases at least Q20
*** NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
*** NOTE: This is a 'working draft' sequence. It currently
consists of 66 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 1015: contig of 1015 bp in length
* 1016 1115: gap of unknown length
* 1116 2278: contig of 1163 bp in length
* 2279 3604: contig of 1226 bp in length
* 3605 3704: gap of unknown length
* 3705 4991: contig of 1287 bp in length
* 4992 5091: gap of unknown length
* 5092 6229: contig of 1138 bp in length
* 6230 6329: gap of unknown length
* 6330 7571: contig of 1242 bp in length
* 7572 7672: contig of unknown length
* 7673 8696: contig of 1024 bp in length
* 8697 8925: gap of unknown length
* 8926 9925: contig of 1030 bp in length
* 9926 11354: gap of unknown length
* 11355 11454: contig of 1429 bp in length
* 11455 13072: gap of unknown length
* 13073 13173: contig of 1618 bp in length
* 13174 14918: contig of 1746 bp in length
* 14919 15018: gap of unknown length
* 15019 16055: contig of 1037 bp in length
* 16056 16155: gap of unknown length
* 16156 17599: contig of 1444 bp in length
* 17600 17699: gap of unknown length
* 17700 19036: contig of 1337 bp in length
* 19037 19136: gap of unknown length
* 19137 20419: contig of 1283 bp in length
* 20420 20519: gap of unknown length
* 20520 21591: contig of 1071 bp in length
* 21592 21691: gap of unknown length
* 21692 23062: contig of 1372 bp in length
* 23063 23162: gap of unknown length
* 23163 24367: contig of 1205 bp in length
* 24368 24467: gap of unknown length
* 24468 25819: contig of 1352 bp in length
* 25820 25919: gap of unknown length
* 25920 27283: contig of 1364 bp in length
* 27284 28628: gap of unknown length
* 28629 28628: contig of 1245 bp in length

```



```

* 28529 28728: gap of unknown length
* 28729 30107: contig of 1379 bp in length
* 30108 30207: gap of unknown length
* 30208 31892: contig of 1685 bp in length
* 31893 31992: gap of unknown length
* 31993 33557: contig of 1565 bp in length
* 33558 33657: gap of unknown length
* 33658 34822: contig of 1165 bp in length
* 34823 34922: gap of unknown length
* 34923 36315: contig of 1393 bp in length
* 36316 36415: gap of unknown length
* 36416 38548: contig of 2133 bp in length
* 38549 38648: gap of unknown length
* 38649 40536: contig of 1888 bp in length
* 40537 40636: gap of unknown length
* 40637 41892: contig of 1256 bp in length
* 41893 41992: gap of unknown length
* 41993 43516: contig of 1524 bp in length
* 43517 43617: gap of unknown length
* 43617 44992: contig of 1376 bp in length
* 44993 45092: gap of unknown length
* 45093 46471: contig of 1379 bp in length
* 46472 46571: gap of unknown length
* 46572 48668: contig of 2097 bp in length
* 48669 48768: gap of unknown length
* 48769 50172: contig of 1404 bp in length
* 50173 50272: gap of unknown length
* 50273 53281: contig of 3009 bp in length
* 53282 53381: gap of unknown length
* 53382 55305: contig of 1924 bp in length
* 55306 55405: gap of unknown length
* 55406 56856: contig of 1451 bp in length
* 56857 56956: gap of unknown length
* 56957 59461: contig of 2405 bp in length
* 59462 59461: gap of unknown length
* 59462 61457: contig of 1996 bp in length
* 61458 61557: gap of unknown length
* 61558 63257: contig of 1700 bp in length
* 63258 63357: gap of unknown length
* 63358 65398: contig of 2041 bp in length
* 65399 65498: gap of unknown length
* 65499 69630: contig of 4132 bp in length
* 69631 69730: gap of unknown length
* 69731 72934: contig of 3204 bp in length
* 72935 73034: gap of unknown length
* 73035 76310: contig of 3276 bp in length
* 7311 76410: gap of unknown length
* 76411 78035: contig of 1625 bp in length
* 78036 78135: gap of unknown length
* 78136 80493: contig of 2358 bp in length
* 80494 80593: gap of unknown length
* 80594 83362: contig of 2769 bp in length
* 83363 83462: gap of unknown length
* 83463 85944: contig of 2382 bp in length
* 85945 85944: gap of unknown length
* 85945 89228: contig of 3284 bp in length
* 89229 89328: gap of unknown length
* 89329 92715: contig of 3387 bp in length
* 92716 92815: gap of unknown length
* 92816 92824: contig of 3409 bp in length
* 92825 96324: gap of unknown length
* 96325 96335: contig of 3309 bp in length
* 96336 99734: gap of unknown length
* 99734 103914: contig of 4181 bp in length
* 103915 104014: gap of unknown length
* 104015 109640: contig of 5626 bp in length
* 109641 109740: gap of unknown length

```

Query Match 0.6%; Score 21; DB 2; Length 189748;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2988 TCAGGAAGTGAAGCACAAG 3008

||||| TCAGGAAGTGAAGCACAAG 97956

# RESULT 49 AC009466

LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

AC009466 193279 bp DNA linear PRI 15-DEC-2000  
 Homo sapiens chromosome 11, clone RP11-87N22, complete sequence.  
 AC009466  
 AC009466.17 GI:11321807  
 HTG.

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 1 (bases 1 to 193279)  
 Homo sapiens chromosome 11, clone RP11-87N22

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 193279)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
 Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,  
 Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,  
 Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,  
 Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,  
 Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,B., Grant,G.,  
 Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,  
 Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,  
 Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,  
 Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,  
 Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,  
 Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,  
 Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,  
 Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.  
 Direct Submission  
 Submitted (24-AUG-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 193279)

TITLE

JOURNAL

REFERENCE

AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
 Barna,N., Bastien,V., Bouckavsky,L., Bouckalter,B., Brown,A.,  
 Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,  
 Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
 Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,  
 Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,  
 Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,  
 McPheeters,R., Meldrum,J., Meneus,L., Mihova,I., Mienga,V.,  
 Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,  
 Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,  
 Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P.,  
 Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Travers,M., Travis,N., Trigglio,J., Vassiliev,H., Viel,R., Vo,A.,  
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,  
 Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (15-DEC-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Nov 23, 2000 this sequence version replaced gi:11181808.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE

JOURNAL

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information -----

Center project name: L1361  
Center clone name: 87\_N\_22

## FEATURES

## Source

## Location/Qualifiers

```

1..193279
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-87N22"
/clone.lib="RP11-11 Human Male BAC"
504..561
/rpt_family="(CA)n"
complement(573..882)
/rpt_family="AluY"
988..1015
/rpt_family="(TG)n"
complement(1077..1265)
/rpt_family="MIR"
complement(1319..1361)
/rpt_family="MIR"
complement(1646..1835)
/rpt_family="MIR"
complement(1943..2320)
/rpt_family="THE1C"
complement(3451..3717)
/rpt_family="AluSg"
3852..3981
/rpt_family="MIR"
4678..4873
/rpt_family="L1PA3"
4907..4926
/rpt_family="(A)n"
5590..5758
/rpt_family="MIR3"
7409..7450
/rpt_family="T-rich"
9037..9083
/rpt_family="L2"
11416..11438
/rpt_family="AT-rich"
13119..13949
/rpt_family="L2"
complement(14968..15042)
/rpt_family="L2"
15058..15122
/rpt_family="L2"
complement(15282..15441)
/rpt_family="MIR"
15848..15899
/rpt_family="(CA)n"
complement(17151..17264)
/rpt_family="MER94"
17719..17739
/rpt_family="AT-rich"
17769..17839
/rpt_family="T-rich"
18599..18853
/rpt_family="L2"
complement(21224..21528)
/rpt_family="AluY"
complement(22197..22307)
/rpt_family="L3"
complement(22450..22875)
/rpt_family="L3"
complement(22884..23008)
/rpt_family="MER20B"
complement(23232..23349)
/rpt_family="L3"
complement(25680..25985)
/rpt_family="AluY"
29025..35170
/rpt_family="L1PA5"

```

```

repeat_region 36356..36375
/rpt_family="(CTG)n"
complement(39370..39845)
/rpt_family="L1MB3"
complement(39876..40070)
/rpt_family="L1M4"
42933..43060
/rpt_family="(TA)n"
43089..43144
/rpt_family="AT-rich"
complement(44577..44696)
/rpt_family="AluYa5/8"
46470..46529
/rpt_family="MIR"
47072..47286
/rpt_family="MIR"
47416..47698
/rpt_family="AluSx"
49722..49765
/rpt_family="(TG)n"
50251..50528
/rpt_family="AluJb"
complement(56872..56178)
/rpt_family="AluSx"
57210..57502
/rpt_family="AluJo"
57596..57678
/rpt_family="L2"
complement(57714..58028)
/rpt_family="AluSx"
58029..58051
/rpt_family="AT-rich"
complement(58542..58602)
/rpt_family="MIR3"
58932..59045
/rpt_family="MER81"
complement(59500..59545)
/rpt_family="MERSA"
60676..60695
/rpt_family="(A)n"
complement(61438..61631)
/rpt_family="MIR"
62350..62580
/rpt_family="MIR"
62788..63827
/rpt_family="AT-rich"
63566..63867
/rpt_family="L1MA4"
63934..64117
/rpt_family="MIR"
complement(64123..64428)
/rpt_family="AluYa5"
complement(64779..65078)
/rpt_family="AluSx"
67738..67813
/rpt_family="CT-rich"
complement(71638..71784)
/rpt_family="MERSA"
complement(71960..72305)
/rpt_family="L3"

```

Query Match 0.6%; Score 21; DB 9; Length 193279;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 579 AATTTTTCCTTTTATCTA 599  
|||||

DB 67736 AATTTTTCCTTTTATCTA 67756  
|||||

RESULT 50  
AC055817  
LOCUS

AC055817

202781 bp

DNA

linear

HTG 06-AUG-2002

```

DEFINITION Mus musculus clone RP23-154I7, *** SEQUENCING IN PROGRESS ***, 8
unordered pieces.
ACCESSION AC035817
VERSION AC035817.5 GI:22123678
KEYWORDS HTG: HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Birtten,B., Nusbaum,C. and Lander,E.
1 (bases 1 to 202781)
Mus musculus, clone RP23-154I7
Unpublished
2 (bases 1 to 202781)
Birtten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kars,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (18-Apr-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 202781)
Birtten,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collimore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 6, 2002 this sequence version replaced gi:14140334.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8049
Center clone name: 154_1_7
-----

```

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

* 1 32467: contig of 32467 bp in length
* 32468 32567: gap of 100 bp
* 32568 66596: contig of 34029 bp in length
* 66597 66696: gap of 100 bp
* 66697 88421: contig of 21725 bp in length
* 88422 88521: gap of 100 bp
* 88522 93034: contig of 4513 bp in length
* 93035 93134: gap of 100 bp
* 93135 131293: contig of 38159 bp in length
* 131294 131393: gap of 100 bp
* 131394 178881: contig of 47488 bp in length
* 178882 178981: gap of 100 bp
* 178982 201438: contig of 22457 bp in length
* 201439 201538: gap of 100 bp
* 201539 202781: contig of 1243 bp in length.

```

#### FEATURES

```

source
1..202781
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-154I7"

```

```

BASE COUNT 56129 a 44477 c 44824 g 56482 t 869 others
ORIGIN

```

```

Query Match 0.6%; Score 21; DB 2; Length 202781;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 2694 TGGACCATGATGTCACAAATG 2714
|||||

```

```

Db 168432 TGGACCATGATGTCACAAATG 168452
|||||

```

```

Search completed: November 6, 2002, 05:43:43
Job time : 22032 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 19:30:52 : Search time 705 Seconds  
(without alignments)  
11918.022 Million cell updates/sec

Title: US-09-782-874-1  
Perfect score: 3731  
Sequence: 1 GAATAATCTTTACTTACTT.....AGTTTCATCTTCTCTCTAAA 3731

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database : N\_Geneseq\_101002.\*  
1: /SID32/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID32/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID32/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID32/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID32/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID32/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID32/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID32/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID32/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID32/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID32/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID32/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID32/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID32/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID32/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID32/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID32/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID32/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID32/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID32/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID32/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID32/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID32/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID32/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3731	100.0	3731	22	Tomato RNA-directe
2	3731	100.0	3731	23	Tomato RNA-directe
3	3495	93.7	3731	23	Tomato RNA-directe
C 4	30	0.8	906	17	Potato calmodulin
C 5	25	0.7	25	22	RACE PCR primer, G
C 6	25	0.7	25	22	RACE PCR primer, G
C 7	25	0.7	25	23	Tomato RNA-directe
C 8	25	0.7	25	23	Tomato RNA-directe
C 9	24	0.6	25	22	RACE PCR primer, G

c 10	24	0.6	25	23	AAS17839	Tomato RNA-directe
c 11	20	0.5	407	24	ABL64204	Stomach cancer rel
c 12	20	0.5	786	24	ABQ55295	Human ovarian anti
c 13	20	0.5	1000	24	AAD38196	Non-typeable Haemo
c 14	20	0.5	1509	20	AAZ52915	Human prostate tum
c 15	20	0.5	1730	20	AAZ18199	TIP49 coding sequ
c 16	20	0.5	1744	21	AAZ59220	Human secreted pro
c 17	20	0.5	1750	20	AAZ24059	Human regulatory m
c 18	20	0.5	38653	22	AAZ72348	Human immune/haema
c 19	20	0.5	34	22	AAD04374	Forward PCR primer
c 20	19	0.5	34	23	AAS17841	Tomato RNA-directe
c 21	19	0.5	255	22	AAH81903	Rat differential t
c 22	19	0.5	327	24	ABN67264	Streptococcus poly
c 23	19	0.5	369	21	AAZ311397	Human secreted pro
c 24	19	0.5	452	22	AA187435	Human polynucleoti
c 25	19	0.5	456	18	AAZ49104	Partial DNA clone
c 26	19	0.5	461	22	ABA59152	Human foetal liver
c 27	19	0.5	461	22	AAK07338	Human brain expres
c 28	19	0.5	461	22	AAK33106	Human bone marrow
c 29	19	0.5	461	22	AAI38901	Probe #7587 used t
c 30	19	0.5	461	22	ABQ07923	Human genome-deriv
c 31	19	0.5	757	22	AAZ22601	Human breast cance
c 32	19	0.5	783	21	AAZ38539	Arabidopsis thalia
c 33	19	0.5	889	22	AAI94184	Human neuroblastom
c 34	19	0.5	1296	20	AAZ05849	Mouse pheromone re
c 35	19	0.5	1405	21	AAA61289	Human secreted pro
c 36	19	0.5	1521	20	AAZ05850	Mouse pheromone re
c 37	19	0.5	2606	23	ABL08786	Drosophila melanog
c 38	19	0.5	2732	20	AAZ05810	Mouse pheromone re
c 39	19	0.5	2962	20	AAZ05811	Mouse pheromone re
c 40	19	0.5	4446	19	AAZ36034	DNA encoding a DNA
c 41	19	0.5	4656	22	AAZ36325	Human musculoskele
c 42	19	0.5	4877	22	AAZ11120	Human small cell l
c 43	19	0.5	4924	21	AAZ75998	Human ORFX ORF1553
c 44	19	0.5	5415	23	ABL30262	Drosophila melanog
c 45	19	0.5	18234	22	AAH62719	Shrimp white spot
c 46	19	0.5	305107	22	AAH62689	Reverse PCR primer
c 47	19	0.5	35	22	AAD04376	Tomato RNA-directe
c 48	18	0.5	35	23	AAS17843	Tomato RNA-directe
c 49	18	0.5	48	22	AAZ29312	Primer base sequen

#### ALIGNMENTS

#### RESULT 1

AAZ04370  
ID AAD04370 standard; cDNA: 3731 BP.

XX AAD04370;

XX AC AAD04370;

XX DT 04-JUL-2001 (first entry)

XX DE Tomato RNA-directed RNA polymerase (RdRP) cDNA.

XX KW Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;

XX KW transgenic plant; tissue culture; plant breeding; therapy; C-protein; ss.

XX OS Lycopersicon esculentum.

XX PH Key Location/Qualifiers

XX FT CDS 194..3538

XX FT /product= "Tomato C-protein having RNA-directed RNA

XX FT /tag= a polymerase (RdRP) activity"

XX PN US6218142-B1.

XX PD 17-APR-2001.

XX PF 05-MAR-1997; 97US-0811583.

XX XX

PR 05-MAR-1997; 97US-0811583.

XX (WASS/) WASENEGGER M.

PA (RIED/) RIEDEL L.

XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;

XX WPI; 2001-289830/30.

XX DR P-PSDB; AAE00897.

XX Claim 1; Column 25-34; 31pp; English.

XX The present sequence is a cDNA encoding tomato C-protein having  
CC RNA-directed RNA polymerase (RdRP) activity. This protein is capable  
CC of RNA-directed RNA synthesis, thus using RNA as a template for  
CC synthesising complementary RNA molecules. RdRP nucleic acid is useful  
CC for modulating gene expression in plants, humans and animals. This may  
CC lead to various physiological, developmental and/or morphological  
CC changes. Transgenic plants containing RdRP nucleic acid is especially  
CC useful in plant cell or tissue cultures and in plant breeding. RdRP is  
CC useful in gene therapy, particularly for treating a disease that is  
CC caused by the undesirable expression or overexpression of a gene.

XX Query Match 100.0%; Score 3731; DB 22; Length 3731;

XX Best Local Similarity 100.0%; Pred. No. 0;

XX Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 GAATATCTCTTACTTACTTACACAGGATTTGACTTACATCACCCTCCCTCAAGCTTTTGT 60

XX 1 GAATATCTCTTACTTACTTACACAGGATTTGACTTACATCACCCTCCCTCAAGCTTTTGT 60

XX 61 GTTGTGATAAATTTGGTGTGCTTCAGTTTCAGTCTACTGCTGGGTAGTTTAT 120

XX 61 GTTGTGATAAATTTGGTGTGCTTCAGTTTCAGTCTACTGCTGGGTAGTTTAT 120

XX 121 TTTCCATACTTCAGGGGATTTCCAGTTGGTGTAGCAATTTGAAGTCGAATCGACT 180

XX 121 TTTCCATACTTCAGGGGATTTCCAGTTGGTGTAGCAATTTGAAGTCGAATCGACT 180

XX 181 GGAATTTGGCTACATGGGAAGACAATTCAGGTTTTCGGATTCCCTTATCTCTG 240

XX 181 GGAATTTGGCTACATGGGAAGACAATTCAGGTTTTCGGATTCCCTTATCTCTG 240

XX 241 GGAAGTGGTTAAGTCATTTAGAGAAATATACAGATATGGAATGATGCAATGGA 300

XX 241 GGAAGTGGTTAAGTCATTTAGAGAAATATACAGATATGGAATGATGCAATGGA 300

XX 301 GGTATAACAGTCCAAAGGAGATCTAGAGATTTGCCAAGTCAATTTGCCGACAAT 360

XX 301 GGTATAACAGTCCAAAGGAGATCTAGAGATTTGCCAAGTCAATTTGCCGACAAT 360

XX 361 AAGTGTGACAAATCATCTTGGCTAATACAGCTGTATTTGGCTCTTCTATT 420

XX 361 AAGTGTGACAAATCATCTTGGCTAATACAGCTGTATTTGGCTCTTCTATT 420

XX 421 GAAGGCTTGGAAATGAATGATATTTGCCAATTCGCGGATATGGAATGATGGA 480

XX 421 GAAGGCTTGGAAATGAATGATATTTGCCAATTCGCGGATATGGAATGATGGA 480

XX 481 TGGCATACATTTGGAATTCGAGTGTAGATATCAGATGACAAGTTGCAAGCTTGGGAG 540

XX 481 TGGCATACATTTGGAATTCGAGTGTAGATATCAGATGACAAGTTGCAAGCTTGGGAG 540

XX 541 TACAGAGTTTCAATTTGGCATTTGGCATTTGGCATTTGGCATTTGGCATTTGGCAT 600

XX 541 TACAGAGTTTCAATTTGGCATTTGGCATTTGGCATTTGGCATTTGGCATTTGGCAT 600

QY 601 TGGTTACAGTACTATAAAGTTCAGCTTTCATATGAAAATATATGCAGGTTGCTCCA 660

Db 601 TGGTTACAGTACTATAAAGTTCAGCTTTCATATGAAAATATATGCAGGTTGCTCCA 660

QY 661 TCGTCCATATGTTCAAAATGCTCAGTTTCTCCTCATACAGTTATTTGGTGTCCCGGAT 720

Db 661 TCGTCCATATGTTCAAAATGCTCAGTTTCTCCTCATACAGTTATTTGGTGTCCCGGAT 720

QY 721 CTATAGAGACTTTGAAACTCCCTGTATAGCTTCTTTAGGAAACTCCCTGATGATCAGTG 780

Db 721 CTATAGAGACTTTGAAACTCCCTGTATAGCTTCTTTAGGAAACTCCCTGATGATCAGTG 780

QY 781 GGTGAGACAAACAGATTTCCCTCCATCTTTGGATAGGGCTATCTTTAGCTTATGTTTGA 840

Db 781 GGTGAGACAAACAGATTTCCCTCCATCTTTGGATAGGGCTATCTTTAGCTTATGTTTGA 840

QY 841 GTTCCCTAGGGGTGCTGCTTCCAAATTTCCAGGAAAGTTTTCCTACATGACAGACG 900

Db 841 GTTCCCTAGGGGTGCTGCTTCCAAATTTCCAGGAAAGTTTTCCTACATGACAGACG 900

QY 901 TGAACAATATTTACTTTACAGACTGGTTTACCTTTTCGCTCTCAAAAATCGGCTCT 960

Db 901 TGAACAATATTTACTTTACAGACTGGTTTACCTTTTCGCTCTCAAAAATCGGCTCT 960

QY 961 GGTCCCAATGTCAGGCTCCGGAAGAAATTTCAATTCCTCCTACAGATTTTGTTCAAAAT 1020

Db 961 GGTCCCAATGTCAGGCTCCGGAAGAAATTTCAATTCCTCCTACAGATTTTGTTCAAAAT 1020

QY 1021 TAGTCTTTGGTACAGCATGGATGCTACCTGGGCGAGCAATTAATGCTACTTTTCG 1080

Db 1021 TAGTCTTTGGTACAGCATGGATGCTACCTGGGCGAGCAATTAATGCTACTTTTCG 1080

QY 1081 ATTAGTTGATCCTCGAAGGAGAAATGTCGATGCTTGGATGCTTACAGAACTGTA 1140

Db 1081 ATTAGTTGATCCTCGAAGGAGAAATGTCGATGCTTGGATGCTTACAGAACTGTA 1140

QY 1141 CTATAAAGAGATGCTGTATGATCCCGTGGTGGCTCAGTACGATGATGATGATGATG 1200

Db 1141 CTATAAAGAGATGCTGTATGATCCCGTGGTGGCTCAGTACGATGATGATGATGATG 1200

QY 1201 TCTCAAGGTPAGCAACCTCCAAATCTCCGTCCTACCTTTAGATGATGATGATGATG 1260

Db 1201 TCTCAAGGTPAGCAACCTCCAAATCTCCGTCCTACCTTTAGATGATGATGATGATG 1260

QY 1261 TGTAAAGAGGTCCTAGTACACCATGCAAGTTTATTTTGTGTCAGAGGTTAATGT 1320

Db 1261 TGTAAAGAGGTCCTAGTACACCATGCAAGTTTATTTTGTGTCAGAGGTTAATGT 1320

QY 1321 TTCCAATCGGTTCTCCGCAATTTATCTGAAGACATAGATAACTTTCTCTGTTTCTTT 1380

Db 1321 TTCCAATCGGTTCTCCGCAATTTATCTGAAGACATAGATAACTTTCTCTGTTTCTTT 1380

QY 1381 TGTGATGAGGATGGGAGAACTGTATTTCTACAGACTTATACCAAGCAAGTACTGG 1440

Db 1381 TGTGATGAGGATGGGAGAACTGTATTTCTACAGACTTATACCAAGCAAGTACTGG 1440

QY 1441 AAGTGTGTCTCAGACAAACATCTATGAGAGGATCTTATCAACTCTCGGAAAGGCTTGT 1500

Db 1441 AAGTGTGTCTCAGACAAACATCTATGAGAGGATCTTATCAACTCTCGGAAAGGCTTGT 1500

QY 1501 AATTGGTGATATAAATTTGAATTTCTTGCATTTTCAATCGAGGCTTGGGATTAATTC 1560

Db 1501 AATTGGTGATATAAATTTGAATTTCTTGCATTTTCAATCGAGGCTTGGGATTAATTC 1560

QY 1561 AGTGTGATGATTTGCAATGAGCTGGCTTACTGCAATGATATAAGAGCTTGGATGGG 1620

Db 1561 AGTGTGATGATTTGCAATGAGCTGGCTTACTGCAATGATATAAGAGCTTGGATGGG 1620

QY 1621 TGAATTTTCCAGATCAAGAAATGTCGAAAATATGCTGCCAGCTTGTTCATCTTTGG 1680

Db 1621 TGAATTTTCCAGATCAAGAAATGTCGAAAATATGCTGCCAGCTTGTTCATCTTTGG 1680

QY 1681 TTCTCCAGAGACATTTGAGTGTCTTATAGGCATGAGATTGAAGTTATTCCCGATGATAA 1740  
DB 1681 TTCTCCAGAGACATTTGAGTGTCTTATAGGCATGAGATTGAAGTTATTCCCGATGATAA 1740  
QY 1741 GGTTCATGGAACACAGCTATGCTCTTTCTGATGGAATTTGTAATAATCTGGTGACTTTGC 1800  
DB 1741 GGTTCATGGAACACAGCTATGCTCTTTCTGATGGAATTTGTAATAATCTGGTGACTTTGC 1800  
QY 1801 TCATAGAGTTGCTCAAAATGTGCGCTTCAATATACCCCATCTGCTTCCAGATTGCTTA 1860  
DB 1801 TCATAGAGTTGCTCAAAATGTGCGCTTCAATATACCCCATCTGCTTCCAGATTGCTTA 1860  
QY 1861 TGTGGATATAAAGGTGTGTTGGGTGTGATCCGGATTCATCAATCAAGTTGCTTTGAG 1920  
DB 1861 TGTGGATATAAAGGTGTGTTGGGTGTGATCCGGATTCATCAATCAAGTTGCTTTGAG 1920  
QY 1921 AAAGAGCATGTCGAATATGAATCAGACAAACATAAAGTTAGATGTCCTTGGATGGAGCAA 1980  
DB 1921 AAAGAGCATGTCGAATATGAATCAGACAAACATAAAGTTAGATGTCCTTGGATGGAGCAA 1980  
QY 1981 ATATCAGCCTTGTAATCTTAATGCTCACTGATTAAGCTTCTCTACACTTGGAGTGA 2040  
DB 1981 ATATCAGCCTTGTAATCTTAATGCTCACTGATTAAGCTTCTCTACACTTGGAGTGA 2040  
QY 2041 AGATGAAGTTCTCGAAGAGCAAGAAAGAGCTGTAGATCAGCTTGATCTATCTTGCA 2100  
DB 2041 AGATGAAGTTCTCGAAGAGCAAGAAAGAGCTGTAGATCAGCTTGATCTATCTTGCA 2100  
QY 2101 TGATCTTTGAAGCAGACAGAGGCTTTGGAATGATGTCCTCGGAGAACACTAATAT 2160  
DB 2101 TGATCTTTGAAGCAGACAGAGGCTTTGGAATGATGTCCTCGGAGAACACTAATAT 2160  
QY 2161 TCTCAAGCCTAATGCTAACTGTGTTATAAGCCTGATGCTGAGCCCTTTCTTCAATGAT 2220  
DB 2161 TCTCAAGCCTAATGCTAACTGTGTTATAAGCCTGATGCTGAGCCCTTTCTTCAATGAT 2220  
QY 2221 GTTGCACACCTTCGCGCATCCCAAGTTGCTGATTTGCGGACTAGATCAAGATATTTAT 2280  
DB 2221 GTTGCACACCTTCGCGCATCCCAAGTTGCTGATTTGCGGACTAGATCAAGATATTTAT 2280  
QY 2281 TCCAAATGGAAGAACATGATGGATGTTGGATGAATCCAGAACCTTGGAAATGGTCA 2340  
DB 2281 TCCAAATGGAAGAACATGATGGATGTTGGATGAATCCAGAACCTTGGAAATGGTCA 2340  
QY 2341 GGTGTTTGTTCAGTTTACTGCTGACATGGAGAGTCTTCTGACGATTTACATCCATT 2400  
DB 2341 GGTGTTTGTTCAGTTTACTGCTGACATGGAGAGTCTTCTGACGATTTACATCCATT 2400  
QY 2401 TAATAACAGAGATCCACACAGTAATTTCACTCTGAAGGAAATCTGCTTGTGCAAA 2460  
DB 2401 TAATAACAGAGATCCACACAGTAATTTCACTCTGAAGGAAATCTGCTTGTGCAAA 2460  
QY 2461 AAATCCATGTTGATCCTGCTGATATTCGTTGTTTAAAGGCTGTAAATGTTGAGCGCT 2520  
DB 2461 AAATCCATGTTGATCCTGCTGATATTCGTTGTTTAAAGGCTGTAAATGTTGAGCGCT 2520  
QY 2521 GCACCATAGTATGTTGTTGTTATCCCTCAGAAAGAAAGACCTCATCCGATGA 2580  
DB 2521 GCACCATAGTATGTTGTTGTTATCCCTCAGAAAGAAAGACCTCATCCGATGA 2580  
QY 2581 ATGTTCTGGAGTGATTTGGATGGGATATCTACTTTGTTGTTGGATCAAGACATGAT 2640  
DB 2581 ATGTTCTGGAGTGATTTGGATGGGATATCTACTTTGTTGTTGGATCAAGACATGAT 2640  
QY 2641 CCCGCCAAGCAAGTCCAGCGGATGGAATATCCTCCAGCACCAGCATACAGTTGGACCA 2700  
DB 2641 CCCGCCAAGCAAGTCCAGCGGATGGAATATCCTCCAGCACCAGCATACAGTTGGACCA 2700  
QY 2701 TGATGTCACAAATTCAGGAAGTTGAGAGTACTTCACCAACTATATTGTAATGACAGTTT 2760  
DB 2701 TGATGTCACAAATTCAGGAAGTTGAGAGTACTTCACCAACTATATTGTAATGACAGTTT 2760  
QY 2761 GGGAAATCATAGCAAAATGCCATGTCGTTATTGACAGACAGAACCTGATATGCCATGAG 2820

DB 2761 GGGAAATCATAGCAAAATGCCATGTCGTTATTGACAGACAGAACCTGATATGGCCATGAG 2820  
QY 2821 TGATCCATGCAAAAACCTTGCTGAGCTCTTTTCAATTTGCACTGGACTTCCAAAGACTGG 2880  
DB 2821 TGATCCATGCAAAAACCTTGCTGAGCTCTTTTCAATTTGCACTGGACTTCCAAAGACTGG 2880  
QY 2881 TGTTCCCGTGAATATACCATCTCAGTTGGCCCTAAAGAAATACCCAGACTTCATGGATAA 2940  
DB 2881 TGTTCCCGTGAATATACCATCTCAGTTGGCCCTAAAGAAATACCCAGACTTCATGGATAA 2940  
QY 2941 GCGGACAGACACAGCTATATCTCAGAAAGAGTTATTGGAAAGCTTTTCAGGAAAGTGAA 3000  
DB 2941 GCGGACAGACACAGCTATATCTCAGAAAGAGTTATTGGAAAGCTTTTCAGGAAAGTGAA 3000  
QY 3001 GGACAAAGCACCTCAGGCTAGCTCTATCCGACCTTCAAGAGATGTTGCAAGAGATC 3060  
DB 3001 GGACAAAGCACCTCAGGCTAGCTCTATCCGACCTTCAAGAGATGTTGCAAGAGATC 3060  
QY 3061 ATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACATTGACGAAGCTTTTGACTA 3120  
DB 3061 ATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACATTGACGAAGCTTTTGACTA 3120  
QY 3121 CAAAACCTGAATATGACAAACAGCTGGTAAATTAATGGACTACTATGGCATAAAAACAGA 3180  
DB 3121 CAAAACCTGAATATGACAAACAGCTGGTAAATTAATGGACTACTATGGCATAAAAACAGA 3180  
QY 3181 GCTGNAATACCTAGTGTGGCTATTAAGAGCATCAAAAACCTTTGACCGCAGAAAAGA 3240  
DB 3181 GCTGNAATACCTAGTGTGGCTATTAAGAGCATCAAAAACCTTTGACCGCAGAAAAGA 3240  
QY 3241 TGCTGAGGCCATTAGTGTGCTGAGGCCCTTGAGGAGGAGGCAAGCCCTGTTCAA 3300  
DB 3241 TGCTGAGGCCATTAGTGTGCTGAGGCCCTTGAGGAGGAGGCAAGCCCTGTTCAA 3300  
QY 3301 GAGGCGTAATGATATAGATGACATGTTACCAAAGCTTCGGCTTGGTACACGTTACATA 3360  
DB 3301 GAGGCGTAATGATATAGATGACATGTTACCAAAGCTTCGGCTTGGTACACGTTACATA 3360  
QY 3361 TCATCCTACATATGGGGTGTGCTCAATCAGGGGTTGAAAGAGCTCAITTCATTAGCTT 3420  
DB 3361 TCATCCTACATATGGGGTGTGCTCAATCAGGGGTTGAAAGAGCTCAITTCATTAGCTT 3420  
QY 3421 TCCCTGGTGTGTTATGACCAGCTAATCCAGATTAGAGAGGACAAAGCAGTAACAGGCC 3480  
DB 3421 TCCCTGGTGTGTTATGACCAGCTAATCCAGATTAGAGAGGACAAAGCAGTAACAGGCC 3480  
QY 3481 AGTTCTCAACTGTCTATCTCTCAGGGCTCACTGAGTCAACAGATTAGTGTGAAATGAGA 3540  
DB 3481 AGTTCTCAACTGTCTATCTCTCAGGGCTCACTGAGTCAACAGATTAGTGTGAAATGAGA 3540  
QY 3541 TTCCAGTCAGCGGTTAAGCTGATATATATGTAATAGGGTGTGATCATAAAGAAAC 3600  
DB 3541 TTCCAGTCAGCGGTTAAGCTGATATATATGTAATAGGGTGTGATCATAAAGAAAC 3600  
QY 3601 TGTATGCACTTGTGACTACCTTTTCTTTAAACTGCATGAAGCTGAACATATATGC 3660  
DB 3601 TGTATGCACTTGTGACTACCTTTTCTTTAAACTGCATGAAGCTGAACATATATGC 3660  
QY 3661 AGTACTCTAAGAAACAGATGTCAGCTAAGTACTAATATGTAATGTAATGATTTGATTTTCATC 3720  
DB 3661 AGTACTCTAAGAAACAGATGTCAGCTAAGTACTAATATGTAATGTAATGATTTGATTTTCATC 3720  
QY 3721 TTTCTTCTAAA 3731  
DB 3721 TTTCTTCTAAA 3731

RESULT 2  
AAS17837  
ID AAS17837 standard; cdna; 3731 BP.  
XX  
AC AAS17837;

XX 08-MAY-2002 (first entry)  
XX  
XX Tomato RNA-directed RNA polymerase (RdRP) cDNA.  
DE  
KW Cytostatic; RdRP; RNA-directed RNA polymerase; in vitro transcription;  
KW Cytostatic; virucide; RNA synthesis inhibitor; antibody; immunogen;  
KW transgenic plant; transgenic animal; cancer; viral infection;  
KW immunoprecipitation; immunolocalisation; ss; gene therapy.  
XX  
OS Lycopersicon esculentum.  
XX  
FH Key Location/Qualifiers  
FH CDS 194..3538  
FT /\*tag= a  
FT /product= "RdRP protein"  
FT /EC\_number= "2.7.7.48"  
XX  
XX US2001023067-A1.  
XX  
XX 20-SEP-2001.  
XX  
XX 08-FEB-2001; 2001US-0782874.  
XX  
XX 05-MAR-1997; 97US-0811583.  
XX  
XX (WASS/) WASSENEGGER M.  
XX (RIED/) RIEDEL L.  
XX  
XX Wassenecker M, Riedel L, Schiebel W, Sanger HL;  
XX WPI; 2001-595798/67.  
XX P-PSDB; AAU10006.  
XX  
XX New nucleic acid molecule encoding a polypeptide having the enzymatic  
XX activity of RNA-directed RNA polymerase, for modulating gene expression  
XX and treating cancer and virus infection in human and animals  
XX  
XX Claim 1; Page 15-19; 34pp; English.  
XX  
XX This sequence represents a cDNA encoding the tomato RNA-directed RNA  
XX polymerase (RdRP) protein of the invention. The invention comprises the  
XX nucleic acid and protein sequences of RdRP. The protein of the invention  
XX can catalyze in vitro transcription of short single stranded RNAs into  
XX DNA molecules, this transcription can be either primed by RNA or DNA  
XX oligonucleotides or be unprimed. The protein may have cytotostatic or  
XX virucide activities. The sequences of the invention may be used in gene  
XX therapy or as an RNA directed RNA synthesis inhibitor. The RdRP cDNA  
XX sequence and a template nucleic acid molecule derived from a nucleic  
XX acid molecule which causes a disease are useful for treating a disease  
XX caused by the undesired expression or overexpression of a nucleic acid  
XX molecule in a human, rat or mouse, by administering the molecules. This  
XX system can be used in the preparation of a pharmaceutical composition  
XX and for inhibiting expression of any desired gene by transferring the  
XX RdRP system to organisms that either lack a comparable mechanism or do  
XX not sufficiently express their own RdRP. An antibody or an antagonist or  
XX inhibitor to the protein are useful for inhibiting RNA directed RNA  
XX synthesis and for ensuring stable heterologous, gene expression in  
XX transgenic organisms. The sequence is useful for probes and/or for the  
XX control of gene expression, as primers for amplification of nucleic acid  
XX molecules and as tools for the detection of expression of the cDNA  
XX molecules. Additionally, nucleotide and protein sequences are useful for  
XX suppression of undesired gene expression in humans and animals. The RdRP  
XX is useful as a therapeutic agent for the control of cancer and virus  
XX infection in humans and animals and the antibody is useful for  
XX immunoprecipitation or immunolocalisation of the protein, identification  
XX of polypeptides interacting with it and screening expression libraries.  
XX  
XX Sequence 3731 BP; 1064 A; 669 C; 849 G; 1149 T; 0 other;  
XX  
XX Query Match 100.0%; Score 3731; DB 23; Length 3731;  
XX Best Local Similarity 100.0%; Pred. No. 0;  
XX Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATATCTCTTACTTACTTCAACCAGGATTTGACTCATCTCCCTCCCTCAAGTCTTTGTGT 60  
DB 1 GAATATCTCTTACTTACTTCAACCAGGATTTGACTCATCTCCCTCCCTCAAGTCTTTGTGT 60  
QY 61 GTTGTGATAATAAATTTGGTTGCTTTCAGTTTCAGTCACTACTGCTGGGTAGTTTTAT 120  
DB 61 GTTGTGATAATAAATTTGGTTGCTTTCAGTTTCAGTCACTACTGCTGGGTAGTTTTAT 120  
QY 121 TTTGCATAACTTCAGGGGTATTCAGTTGGTGTAGCAATTTGAAAGTTCGAACATGCACTT 180  
DB 121 TTTGCATAACTTCAGGGGTATTCAGTTGGTGTAGCAATTTGAAAGTTCGAACATGCACTT 180  
QY 181 GGAATTTGGCTACATGGGAAAGACAATTCAGTTTTCGGATTCCTTATCTCTCTGTC 240  
DB 181 GGAATTTGGCTACATGGGAAAGACAATTCAGTTTTCGGATTCCTTATCTCTCTGTC 240  
QY 241 GGAAGTGTAAAGTCAATTCCTTAGAGAAATATACAGGATATGGAATGTATGTGCAATGGA 300  
DB 241 GGAAGTGTAAAGTCAATTCCTTAGAGAAATATACAGGATATGGAATGTATGTGCAATGGA 300  
QY 301 GGTAAACAGTCCAAAGGAGGATCTAGAGCAATTTGCCAAAGTTCGAATTCGGGACACAT 360  
DB 301 GGTAAACAGTCCAAAGGAGGATCTAGAGCAATTTGCCAAAGTTCGAATTCGGGACACAT 360  
QY 361 AAGTCTGACAAATCATCTTTGGCTAATAACAGGCTGTATTTGGCTCTCTCTTATTT 420  
DB 361 AAGTCTGACAAATCATCTTTGGCTAATAACAGGCTGTATTTGGCTCTCTCTTATTT 420  
QY 421 GAAAGCTTTGGGAAATGAAACTGATATTTGCCAACTGGGGCATATGTGGATCAGATGGA 480  
DB 421 GAAAGCTTTGGGAAATGAAACTGATATTTGCCAACTGGGGCATATGTGGATCAGATGGA 480  
QY 481 TGGGATAACTTTGAATTTCCGATGTCAGATATCAGATCACAAGTTTCAGTTGGGGAAG 540  
DB 481 TGGGATAACTTTGAATTTCCGATGTCAGATATCAGATCACAAGTTTCAGTTGGGGAAG 540  
QY 541 TACAGAAGTTCAATTTCAATTTGGCATTTGGATTTGAAGAAATTTTTTTCTTTTATCTAG 600  
DB 541 TACAGAAGTTCAATTTCAATTTGGCATTTGGATTTGAAGAAATTTTTTTCTTTTATCTAG 600  
QY 601 TGGTTCACTGACTATAAATTTCACTTTTATAGCTTCTTTAAGGAACTCTCTGATGATG 660  
DB 601 TGGTTCACTGACTATAAATTTCACTTTTATAGCTTCTTTAAGGAACTCTCTGATGATG 660  
QY 661 TCGTCCATATGTCATAAATGCTTCAGTTTCTCTCATACAGTATTTGGTCTCTCTCGAT 720  
DB 661 TCGTCCATATGTCATAAATGCTTCAGTTTCTCTCATACAGTATTTGGTCTCTCTCGAT 720  
QY 721 CTATAGAGACTTGAAACTCTCTGTTTATAGCTTCTTTAAGGAACTCTCTGATGATG 780  
DB 721 CTATAGAGACTTGAAACTCTCTGTTTATAGCTTCTTTAAGGAACTCTCTGATGATG 780  
QY 781 GGTGAGGACAACAGATTTCCCTCCATCTTGGATAGGGCTATCTTCTAGCTATGTTTCCA 840  
DB 781 GGTGAGGACAACAGATTTCCCTCCATCTTGGATAGGGCTATCTTCTAGCTATGTTTCCA 840  
QY 841 GTTCGCTAGGGGTGTCGTTCTCCAAATTTCCAGGAAAGTTTTTCCACTATGCAAGACG 900  
DB 841 GTTCGCTAGGGGTGTCGTTCTCCAAATTTCCAGGAAAGTTTTTCCACTATGCAAGACG 900  
QY 901 TGAAACAAATATTTACATTTTACAGCTGGTTTACCTTTTTCGCTCTCAAAATCGGCTCT 960  
DB 901 TGAAACAAATATTTACATTTTACAGCTGGTTTACCTTTTTCGCTCTCAAAATCGGCTCT 960  
QY 961 GGTTCCTCAATGTCCAGGCTCCGGAAGGAATTTCAATTCCTTACAGATTTTCTTCAAAAT 1020  
DB 961 GGTTCCTCAATGTCCAGGCTCCGGAAGGAATTTCAATTCCTTACAGATTTTCTTCAAAAT 1020  
QY 1021 TAGTTCTTTGGTACAGATGGAATGATACCTGGCCAGCATTAATTAATGCTACTTTTCCG 1080  
DB 1021 TAGTTCTTTGGTACAGATGGAATGATACCTGGCCAGCATTAATTAATGCTACTTTTCCG 1080

QY 1081 ATTAGTTGATCTCGAAGGAAATGTGGCATGTGATTCAGCATGCCCTAGAGAACTGTA 1140  
DB 1081 ATTAGTTGATCTCGAAGGAAATGTGGCATGTGATTCAGCATGCCCTAGAGAACTGTA 1140  
QY 1141 CTATATAAAGGAGTCTGTTATGATCCCGTGGGTGGCTCACTCAGCAGATGATGGGTA 1200  
DB 1141 CTATATAAAGGAGTCTGTTATGATCCCGTGGGTGGCTCACTCAGCAGATGATGGGTA 1200  
QY 1201 TCTCAAGGGTAGACAACCTCCAAATCTCCGTCCATCACITTAGATGATGGTGGTGA 1260  
DB 1201 TCTCAAGGGTAGACAACCTCCAAATCTCCGTCCATCACITTAGATGATGGTGGTGA 1260  
QY 1261 TGTAAAGGGTCTAGTAACACCACTGCAAAAGTTTATTTTGTGGTCCAGAGGTTAATGT 1320  
DB 1261 TGTAAAGGGTCTAGTAACACCACTGCAAAAGTTTATTTTGTGGTCCAGAGGTTAATGT 1320  
QY 1321 TTCCAAATCGGGTCTCCGCAATTTCTGAAGACATAGATAACTTTCTTCGTGTTCTTT 1380  
DB 1321 TTCCAAATCGGGTCTCCGCAATTTCTGAAGACATAGATAACTTTCTTCGTGTTCTTT 1380  
QY 1381 TGTGTAGAGAGTGGGAGAACTGTATTCTACAGACTTATTACCAAAACAGTACTGG 1440  
DB 1381 TGTGTAGAGAGTGGGAGAACTGTATTCTACAGACTTATTACCAAAACAGTACTGG 1440  
QY 1441 AAGTGGTGTGAGGACAAACATCTATGAGAGATCTTATCAACTCTGCGGAAAGCTTTGT 1500  
DB 1441 AAGTGGTGTGAGGACAAACATCTATGAGAGATCTTATCAACTCTGCGGAAAGCTTTGT 1500  
QY 1501 AATTGGTGATAAAAAATTTGAATTTCTTCATTTCTCAGCCAGCTTGGGGATTAATTC 1560  
DB 1501 AATTGGTGATAAAAAATTTGAATTTCTTCATTTCTCAGCCAGCTTGGGGATTAATTC 1560  
QY 1561 AGTGTGATGTTGATCAAGAACTGACCTTACTGCAAAATGATTAAGAGCTTGGATGGG 1620  
DB 1561 AGTGTGATGTTGATCAAGAACTGACCTTACTGCAAAATGATTAAGAGCTTGGATGGG 1620  
QY 1621 TGATTTTCCAGATCAAGATCTGCAAAATGATGCTGCCAGACTTGGTCAATCTTTTGG 1680  
DB 1621 TGATTTTCCAGATCAAGATCTGCAAAATGATGCTGCCAGACTTGGTCAATCTTTTGG 1680  
QY 1681 TTCCTCCAGAGAGACTTGGTGTCTTAGGCATGAGATTGAAGTTATCCCGATGATAA 1740  
DB 1681 TTCCTCCAGAGAGACTTGGTGTCTTAGGCATGAGATTGAAGTTATCCCGATGATAA 1740  
QY 1741 GGTTATGGAACAGCTATCTCTTCTGATGGAATGTTGTAATATCTGGTGACTTTTC 1800  
DB 1741 GGTTATGGAACAGCTATCTCTTCTGATGGAATGTTGTAATATCTGGTGACTTTTC 1800  
QY 1801 TCATAGAGTTGCTCAAAATGTGGCTTCAATATACCCATCTGCTTCCAGATTCTGTTA 1860  
DB 1801 TCATAGAGTTGCTCAAAATGTGGCTTCAATATACCCATCTGCTTCCAGATTCTGTTA 1860  
QY 1861 TGTGGATATAAGGTTGTTGGGTGTTGATCCGGATTTCATCAATGAAGTTGTCTTTGAG 1920  
DB 1861 TGTGGATATAAGGTTGTTGGGTGTTGATCCGGATTTCATCAATGAAGTTGTCTTTGAG 1920  
QY 1921 AAGAGCATCTGAAATATGAATCAGACACATAAAGTTAGATGCTCTGGATGGAGCAA 1980  
DB 1921 AAGAGCATCTGAAATATGAATCAGACACATAAAGTTAGATGCTCTGGATGGAGCAA 1980  
QY 1981 ATATCAGCTTGTATCTTAATCGTCAACTGATTAACGCTTGTCTACACTTGGAGTAA 2040  
DB 1981 ATATCAGCTTGTATCTTAATCGTCAACTGATTAACGCTTGTCTACACTTGGAGTAA 2040  
QY 2041 AGATGAAGTTCTCGAAGAGCAAGCAAGAGGAGCTGTAGATCAGCTTGATCTTGTGA 2100  
DB 2041 AGATGAAGTTCTCGAAGAGCAAGCAAGAGGAGCTGTAGATCAGCTTGATCTTGTGA 2100  
QY 2101 TGATTTCTTGAAGGCACAGAGGCTTTGGAATGATGCTCTCGGAGAGAACTAATAT 2160  
DB 2101 TGATTTCTTGAAGGCACAGAGGCTTTGGAATGATGCTCTCGGAGAGAACTAATAT 2160  
QY 2161 TCTCAAGGCAATCTAAACTGTGGTTATAAGCCTGATGCTGAGCCCTTCTTTCAATGAT 2220

DB 2161 TCTCAAGGCAATCTAAACTGTGGTTATAAGCCTGATGCTGAGCCCTTCTTTCAATGAT 2220  
QY 2221 GTTTCGAAACCTTCGCGCATCCCAAGTTGCTCGATTTGCGGACTAGATCAAGATATTTAT 2280  
DB 2221 GTTTCGAAACCTTCGCGCATCCCAAGTTGCTCGATTTGCGGACTAGATCAAGATATTTAT 2280  
QY 2281 TCCAAATGGAAGAACAAATGATGGGATGTTTGGATGAATCCAGAACCTTGGATATGGTCA 2340  
DB 2281 TCCAAATGGAAGAACAAATGATGGGATGTTTGGATGAATCCAGAACCTTGGATATGGTCA 2340  
QY 2341 GGTGTTTGTTCAGTTTACTGCTGACATGGAGAGTTTCTGACGATTACATCCATT 2400  
DB 2341 GGTGTTTGTTCAGTTTACTGCTGACATGGAGAGTTTCTGACGATTACATCCATT 2400  
QY 2401 TAATAACAGACATCCCAACACAGTAATTTCAATCTGAAGGAAATGTGTTGTTGCCAAA 2460  
DB 2401 TAATAACAGACATCCCAACACAGTAATTTCAATCTGAAGGAAATGTGTTGTTGCCAAA 2460  
QY 2461 AATCCATGCTTCATCCTGGTGTATTCGTGTTTAAAGGCTGTAATGTTGAGCGGT 2520  
DB 2461 AATCCATGCTTCATCCTGGTGTATTCGTGTTTAAAGGCTGTAATGTTGAGCGGT 2520  
QY 2521 GCACCATGTTGATGTTGTTGATTCCTCAGAAAGGAAAGACCTCATCCGAATGA 2580  
DB 2521 GCACCATGTTGATGTTGTTGATTCCTCAGAAAGGAAAGACCTCATCCGAATGA 2580  
QY 2581 ATGTTCTGGAGTGATTTGGATGGGATATCTACTTGTGCTGGGATCAAGACATGAT 2640  
DB 2581 ATGTTCTGGAGTGATTTGGATGGGATATCTACTTGTGCTGGGATCAAGACATGAT 2640  
QY 2641 CCGCCAAAGGCAAGTCCAGCCGATGGAATATCCTCAGACCCAGCATACAGTTGGACCA 2700  
DB 2641 CCGCCAAAGGCAAGTCCAGCCGATGGAATATCCTCAGACCCAGCATACAGTTGGACCA 2700  
QY 2701 TGATGTCACAAATGAGGAAGTTGAAGAGTACTTCCACCACTATATGTTGAATGACAGTTT 2760  
DB 2701 TGATGTCACAAATGAGGAAGTTGAAGAGTACTTCCACCACTATATGTTGAATGACAGTTT 2760  
QY 2761 GGAATCATAGCAAAATGCCATGTCGTATTTGACAGACAGAACCTGATATGCCATGAG 2820  
DB 2761 GGAATCATAGCAAAATGCCATGTCGTATTTGACAGACAGAACCTGATATGCCATGAG 2820  
QY 2821 TGATCCATGCAAAACCTTGTGAGTCTTTTCAATTCAGTGGACTTTCCAAAGACTGG 2880  
DB 2821 TGATCCATGCAAAACCTTGTGAGTCTTTTCAATTCAGTGGACTTTCCAAAGACTGG 2880  
QY 2881 TGTTCGCGTGAATACCATCTCAGTTCGCCCTTAAAGATACCCAGACTTCATGGATAA 2940  
DB 2881 TGTTCGCGTGAATACCATCTCAGTTCGCCCTTAAAGATACCCAGACTTCATGGATAA 2940  
QY 2941 GCCGACACAGACCACTATCTCAGAAAGAGTTATTTGAAAGCTTTTCAGGAAAGTGA 3000  
DB 2941 GCCGACACAGACCACTATCTCAGAAAGAGTTATTTGAAAGCTTTTCAGGAAAGTGA 3000  
QY 3001 GGACAAAGCACTCAGCTAGCTCTATCGGACCTTCACAGAGATGTTGCAAGAGATC 3060  
DB 3001 GGACAAAGCACTCAGCTAGCTCTATCGGACCTTCACAGAGATGTTGCAAGAGATC 3060  
QY 3061 ATATGATGCTGATATGGAAGTTGATGGAATTTGAAGATTACATTGACGAGCTTTTGACTA 3120  
DB 3061 ATATGATGCTGATATGGAAGTTGATGGAATTTGAAGATTACATTGACGAGCTTTTGACTA 3120  
QY 3121 CAAAACCTGATATGACAAAGCTGGGTAAATTAATGACTACTATGCACTAAACAGAGA 3180  
DB 3121 CAAAACCTGATATGACAAAGCTGGGTAAATTAATGACTACTATGCACTAAACAGAGA 3180  
QY 3181 GGCTGAAATACTTGTGCTGGCATTATGAAGCATCAAAAACCTTTTGACCCGAGAAAAGA 3240  
DB 3181 GGCTGAAATACTTGTGCTGGCATTATGAAGCATCAAAAACCTTTTGACCCGAGAAAAGA 3240  
QY 3241 TGCTGAGGCCAATAGTGTGTTGAGGGCTTGAAGAGGAGGCAAGAGCTGGTTCAA 3300







```
QY 2493 TTTTAAAGGCTGAAATGTTTCGAGCGCTGCACCACATGGTAGATTGTTGTTATCCCTC 2552
Db 2493 TTTTAAAGGCTGAAATGTTTCGAGCGCTGCACCACATGGTAGATTGTTGTTATCCCTC 2552
QY 2553 AGAAAGGAAAGAACCTCATCCGAATGAATGTTCTGGGAGTGAATGGATGGGATATCT 2612
Db 2553 AGAAAGGAAAGAACCTCATCCGAATGAATGTTCTGGGAGTGAATGGGATGGGATATCT 2612
QY 2613 ACTTTGTTTCTGGGATCAAGACATGATCCGCCCAAGGCAAGTCCAGCGGATGGAATATC 2672
Db 2613 ACTTTGTTTCTGGGATCAAGACATGATCCGCCCAAGGCAAGTCCAGCGGATGGAATATC 2672
QY 2673 CTCACGACCCAGCATACAGTTGGACCATGATGTCACAAATTGAGNAGTTGAAGAGTACT 2732
Db 2673 CTCACGACCCAGCATACAGTTGGACCATGATGTCACAAATTGAGNAGTTGAAGAGTACT 2732
QY 2733 TCACCAACTATATTGTAATGACAGTTTGGGAATCATAGCAAAATGCCCATGTCGTATTG 2792
Db 2733 TCACCAACTATATTGTAATGACAGTTTGGGAATCATAGCAAAATGCCCATGTCGTATTG 2792
QY 2793 CAGACAGAGNACCTGATATGGCCATGAGTATCCATGCAAAAACCTTGTGAGCTCTTTT 2852
Db 2793 CAGACAGAGNACCTGATATGGCCATGAGTATCCATGCAAAAACCTTGTGAGCTCTTTT 2852
QY 2853 CAATTGCAGTGGACTTTCCAAAGACTGGTGTCCCGCTGAAATACCATCTCAGTTGCGGC 2912
Db 2853 CAATTGCAGTGGACTTTCCAAAGACTGGTGTCCCGCTGAAATACCATCTCAGTTGCGGC 2912
QY 2913 CTAAGAATAACCCAGACTTCAATGGAATAAGCGGACAAAGACCATATATCTCAGAAAGAG 2972
Db 2913 CTAAGAATAACCCAGACTTCAATGGAATAAGCGGACAAAGACCATATATCTCAGAAAGAG 2972
QY 2973 TTATTGGAAGCTTTTCAGAAAGTGAAGCAAGCAAAACCTCAGGCTAGCTCTATCGCGA 3032
Db 2973 TTATTGGAAGCTTTTCAGAAAGTGAAGCAAGCAAAACCTCAGGCTAGCTCTATCGCGA 3032
QY 3033 CCTTCAAGAGATGTTGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGGATTG 3092
Db 3033 CCTTCAAGAGATGTTGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGGATTG 3092
QY 3093 AAGATTACATTGACGAGCTTTGACTACAAACCTGAAATATGACACAACTGGGTAAAT 3152
Db 3093 AAGATTACATTGACGAGCTTTGACTACAAACCTGAAATATGACACAACTGGGTAAAT 3152
QY 3153 TAATGGACTACTATGGCATAAACACAGAGCTGAAATACCTTAGTGTGGCATTATGAAG 3212
Db 3153 TAATGGACTACTATGGCATAAACACAGAGCTGAAATACCTTAGTGTGGCATTATGAAG 3212
QY 3213 CATCAAAACCTTTTGACCCGAGAAAGATGCTGAGGCCATTAGTGTGTGAGGGCCT 3272
Db 3213 CATCAAAACCTTTTGACCCGAGAAAGATGCTGAGGCCATTAGTGTGTGAGGGCCT 3272
QY 3273 TGAGGAGGAGGCAAGAGCTGTTTCAAGAGCGTAAATGATATAGATGACATGTTACCAA 3332
Db 3273 TGAGGAGGAGGCAAGAGCTGTTTCAAGAGCGTAAATGATATAGATGACATGTTACCAA 3332
QY 3333 AGGCTTCGGCTTGGTACCAGCTTACATATCATCTACATATTGGGGTGTCTACATCAGG 3392
Db 3333 AGGCTTCGGCTTGGTACCAGCTTACATATCATCTACATATTGGGGTGTCTACATCAGG 3392
QY 3393 GGTGAAAAGAGCTCATTTCAATAGCTTCCCTGGTGTCTTATGACGAGCTAATCCAGA 3452
Db 3393 GGTGAAAAGAGCTCATTTCAATAGCTTCCCTGGTGTCTTATGACGAGCTAATCCAGA 3452
QY 3453 TTAAGAGGACAAAGCACGCTAACAGGCCAGTCTCAACTTGTCACTCTCFCAGGGCTCAAC 3512
Db 3453 TTAAGAGGACAAAGCACGCTAACAGGCCAGTCTCAACTTGTCACTCTCFCAGGGCTCAAC 3512
QY 3513 TGAGTCACAGATTAGTGTGAAATGAGATTCAGTCGAGCGTTAAGCTCATATATATA 3572
Db 3513 TGAGTCACAGATTAGTGTGAAATGAGATTCAGTCGAGCGTTAAGCTCATATATATA 3572
```

```
QY 3573 ATGTAATAGGCTGATCATTAAGAAAACCTGTTATGCATTTGTGACTACCTTTTCTTTA 3632
Db 3573 ATGTAATAGGCTGATCATTAAGAGATCTGTTATGCATTTGTGACTACCTTTTCTTTA 3632
QY 3633 AAACCTGCATGAAGCTGCAACATATATCAGTACTCTAAGAAACAGATGTACAGCTAAGTA 3692
Db 3633 AAACCTGCATGAAGCTGCAACATATATCAGTACTCTAAGAAACAGATGTACAGCTAAGTA 3692
QY 3693 CTAATATATGATGATTGAGTTGAGTTTCATCTTCTCTTAA 3731
Db 3693 CTAATATGATGATTGAGTTGAGTTTCATCTTCTCTTAA 3731

RESULT 4
AAT18078/c
ID AAT18078 standard; cDNA; 906 BP.
XX AAT18078;
XX 17-AUG-1996 (first entry)
XX Potato calmodulin gene.
DE Potato; calmodulin; gene expression; shoot; tuber; growth;
KW development; antisense; increase; decrease; ds.
XX Solanum tuberosum (Isolate p-PCM-1).
FH Key Location/Qualifiers
FT CDS 81..530
FT FT /*tag= a
FT FT /product= Calmodulin.
XX US5498533-A.
XX 12-MAR-1996.
XX 30-JUL-1993; 93US-0100874.
XX 30-JUL-1993; 93US-0100874.
XX (UNIW ) UNIV WASHINGTON STATE RES FOUND.
XX An GH, Han T, Poovalah BW, Takezawa D;
XX WPI; 1996-159687/16.
XX P-PSDB; AAR81571.
XX Transgenic potato plants with increased stem and tuber elongation -
XX contain nucleic acid construct pPCM-1 comprising potato calmodulin
XX coding sequence in the sense direction
XX Claim 6; Columns 7-8; 14pp; English.
XX Controlling growth and development of potato plants can be achieved
XX by increasing or decreasing the expression of a gene encoding
XX calmodulin. Transgenic potato plants carrying sense nucleic acid
XX constructs of p-PCM-1, a cDNA clone of potato calmodulin, exhibit
XX increased shoot and tuber growth yet plants carrying antisense
XX nucleic acid constructs exhibit decreased shoot and tuber growth.
XX Sequence 906 BP; 254 A; 151 C; 219 G; 282 T; 0 other;

Query Match 0.8%; Score 30; DB 17; Length 906;
Best Local Similarity 100.0%; Pred. No. 0.00031;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3687 TAAGTACTAATATGATGATGATTTGAGTTT 3716
Db 47 TAAGTACTAATATGATGATTTGAGTTT 18

RESULT 5
```

```

AAD04371/c
ID AAD04371 standard; DNA; 25 BP.
XX
AC AAD04371;
XX
DT 04-JUL-2001 (first entry)
XX
DE RACE PCR primer, GSP400 to amplify tomato RdRP cDNA.
XX
KW Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
KW transgenic plant; tissue culture; plant breeding; therapy; C-protein;
KW PCR primer; RACE; ss.
XX
OS Lycopersicon esculentum.
XX
PN US6218142-B1.
XX
PD 17-APR-2001.
XX
PF 05-MAR-1997; 97US-0811583.
XX
PR 05-MAR-1997; 97US-0811583.
XX
PA (WASS/) WASSENEGGER M.
PA (RIED/) RIEDEL L.
XX
PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX
DR WPI; 2001-289830/30.
XX
PT New nucleic acid molecules encoding polypeptides with RNA-directed RNA
PT polymerase enzymatic activity, useful in modulating gene expression in
PT plants, humans and animals, as well as in plant cell/tissue cultures or
PT plant breeding -
XX
PS Example 3; Column 20; 31pp; English.
XX
CC The present sequence is RACE (rapid amplification of 5'-cDNA ends) PCR
CC gene specific reverse (GSP) primer, GSP400, used to amplify a cDNA
CC encoding tomato C-protein having RNA-directed RNA polymerase (RdRP)
CC activity. This protein is capable of RNA-directed RNA synthesis, thus
CC using RNA as a template for synthesising complementary RNA molecules.
CC RdRP nucleic acid is useful for modulating gene expression in plants,
CC humans and animals. This may lead to various physiological, developmental
CC and/or morphological changes. Transgenic plants containing RdRP nucleic
CC acid is especially useful in plant cell or tissue cultures and in plant
CC breeding. RdRP is useful in gene therapy, particularly for treating a
CC disease that is caused by the undesirable expression or overexpression of
CC a gene.
XX
SQ Sequence 25 BP; 10 A; 4 C; 7 G; 4 T; 0 other;

Query Match 0.7%; Score 25; DB 22; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.093;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1838 CCATCTGCTTCCAGATTGCTTATG 1862
Db 25 CCATCTGCTTCCAGATTGCTTATG 1

RESULT 6
AAD04373/c
ID AAD04373 standard; DNA; 25 BP.
XX
AC AAD04373;
XX
DT 04-JUL-2001 (first entry)
XX
DE RACE PCR primer, GSP1200 to amplify tomato RdRP cDNA.
XX
KW Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
KW transgenic plant; tissue culture; plant breeding; therapy; C-protein;

```

```

KW PCR primer; RACE; ss.
XX
OS Lycopersicon esculentum.
XX
PN US6218142-B1.
XX
PD 17-APR-2001.
XX
PF 05-MAR-1997; 97US-0811583.
XX
PR 05-MAR-1997; 97US-0811583.
XX
PA (WASS/) WASSENEGGER M.
PA (RIED/) RIEDEL L.
XX
PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX
DR WPI; 2001-289830/30.
XX
PT New nucleic acid molecules encoding polypeptides with RNA-directed RNA
PT polymerase enzymatic activity, useful in modulating gene expression in
PT plants, humans and animals, as well as in plant cell/tissue cultures or
PT plant breeding -
XX
PS Example 3; Column 20; 31pp; English.
XX
CC The present sequence is RACE (rapid amplification of 5'-cDNA ends) PCR
CC gene specific reverse (GSP) primer, GSP1200, used to amplify a cDNA
CC encoding tomato C-protein having RNA-directed RNA polymerase (RdRP)
CC activity. This protein is capable of RNA-directed RNA synthesis, thus
CC using RNA as a template for synthesising complementary RNA molecules.
CC RdRP nucleic acid is useful for modulating gene expression in plants,
CC humans and animals. This may lead to various physiological, developmental
CC and/or morphological changes. Transgenic plants containing RdRP nucleic
CC acid is especially useful in plant cell or tissue cultures and in plant
CC breeding. RdRP is useful in gene therapy, particularly for treating a
CC disease that is caused by the undesirable expression or overexpression of
CC a gene.
XX
SQ Sequence 25 BP; 4 A; 5 C; 10 G; 6 T; 0 other;

Query Match 0.7%; Score 25; DB 22; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.093;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2659 GCCGATGGAAATATCTCTCCAGCACCC 2683
Db 25 GCCGATGGAAATATCTCTCCAGCACCC 1

RESULT 7
AAS17838/c
ID AAS17838 standard; DNA; 25 BP.
XX
AC AAS17838;
XX
DT 08-MAY-2002 (first entry)
XX
DE Tomato RNA-directed RNA polymerase (RdRP) GSP400 PCR primer.
XX
KW Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
KW cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;
KW transgenic plant; transgenic animal; cancer; viral infection; RACE;
KW immunoprecipitation; immunolocalisation; PCR primer; GSP400;
KW gene therapy; ss.
XX
OS Lycopersicon esculentum.
XX
PN US2001023067-A1.
XX
PD 20-SEP-2001.
XX
PF 08-FEB-2001; 2001US-0782874.

```

```

XX PR 05-MAR-1997; 97US-0811583.
XX PA (WASS/) WASSENEGGER M.
XX PA (RIED/) RIEDEL L.
XX PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX DR WPI; 2001-595798/67.
XX PT New nucleic acid molecule encoding a polypeptide having the enzymatic
XX PT activity of RNA-directed RNA polymerase, for modulating gene expression
XX PT and treating cancer and virus infection in human and animals
XX PS Example 3; Page 11; 34pp; English.
XX CC This sequence represents the tomato RNA-directed RNA polymerase (RdRp)
XX CC specific PCR primer GSP400 used to obtain the 5' end of the cDNA
XX CC molecule using the rapid amplification of cDNA ends (RACE) technique.
XX CC The invention comprises the nucleic acid and protein sequences of RdRp,
XX CC the protein of the invention can catalyse in vitro transcription of
XX CC short single stranded RNAs into DNA molecules, this transcription can be
XX CC either primed by RNA or DNA oligonucleotides or be unprimed. The protein
XX CC may have cytosstatic or virucide activities. The sequences of the
XX CC synthesis inhibitor. The RdRp cDNA sequence and a template nucleic acid
XX CC molecule derived from a nucleic acid molecule which causes a disease are
XX CC useful for treating a disease caused by the undesired expression or
XX CC overexpression of a nucleic acid molecule in a human, rat or mouse, by
XX CC administering the molecules. This system can be used in the preparation
XX CC of a pharmaceutical composition and for inhibiting expression of any
XX CC desired gene by transferring the RdRp system to organisms that either
XX CC lack a comparable mechanism or do not sufficiently express their own
XX CC RdRp. An antibody or an antagonist or inhibitor to the protein are
XX CC useful for inhibiting RNA directed RNA synthesis and for ensuring stable
XX CC heterologous, gene expression in transgenic organisms. The sequence is
XX CC useful for probes and/or for the control of gene expression, as primers
XX CC for amplification of nucleic acid molecules and as tools for the
XX CC detection of expression of the cDNA molecules. Additionally, nucleotide
XX CC and protein sequences are useful for suppression of undesired gene
XX CC expression in humans and animals. The RdRp is useful as a therapeutic
XX CC agent for the control of cancer and virus infection in humans and
XX CC animals and the antibody is useful for immunoprecipitation or
XX CC immunolocalisation of the protein, identification of polypeptides
XX CC interacting with it and screening expression libraries.
XX SQ Sequence 25 BP; 10 A; 4 C; 7 G; 4 T; 0 other;

Query Match 0.7%; Score 25; DB 23; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.093;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1838 CCATCTGCTTCCAGATTCGTTATG 1862
DB 25 CCATCTGCTTCCAGATTCGTTATG 1

RESULT 8
AAAS17840/c
ID AAS17840 standard; DNA; 25 BP.
XX AC AAS17840;
XX DT 08-MAY-2002 (first entry)
XX DE Tomato RNA-directed RNA polymerase (RdRp) GSP1200 PCR primer.
XX KW Tomato; RdRp; RNA-directed RNA polymerase; in vitro transcription;
XX KW cytosstatic; virucide; RNA synthesis inhibitor; antibody; immunogen;
XX KW transgenic plant; transgenic animal; cancer; viral infection; RACE;
XX KW immunoprecipitation; immunolocalisation; PCR primer; GSP1200;
XX KW gene therapy; ss.

Lycopersicon esculentum.
US2001023067-A1.
20-SEP-2001.
08-FEB-2001; 2001US-0782874.
05-MAR-1997; 97US-0811583.
(WASS/) WASSENEGGER M.
(RIED/) RIEDEL L.
Wassenegger M, Riedel L, Schiebel W, Sanger HL;
WPI; 2001-595798/67.
New nucleic acid molecule encoding a polypeptide having the enzymatic
activity of RNA-directed RNA polymerase, for modulating gene expression
and treating cancer and virus infection in human and animals
Example 3; Page 11; 34pp; English.
This sequence represents the tomato RNA-directed RNA polymerase (RdRp)
specific PCR primer GSP1200 used to obtain the 5' end of the cDNA
molecule using the rapid amplification of cDNA ends (RACE) technique.
The invention comprises the nucleic acid and protein sequences of RdRp,
the protein of the invention can catalyse in vitro transcription of
short single stranded RNAs into DNA molecules, this transcription can be
either primed by RNA or DNA oligonucleotides or be unprimed. The protein
may have cytosstatic or virucide activities. The sequences of the
synthesis inhibitor. The RdRp cDNA sequence and a template nucleic acid
molecule derived from a nucleic acid molecule which causes a disease are
useful for treating a disease caused by the undesired expression or
overexpression of a nucleic acid molecule in a human, rat or mouse, by
administering the molecules. This system can be used in the preparation
of a pharmaceutical composition and for inhibiting expression of any
desired gene by transferring the RdRp system to organisms that either
lack a comparable mechanism or do not sufficiently express their own
RdRp. An antibody or an antagonist or inhibitor to the protein are
useful for inhibiting RNA directed RNA synthesis and for ensuring stable
heterologous, gene expression in transgenic organisms. The sequence is
useful for probes and/or for the control of gene expression, as primers
for amplification of nucleic acid molecules and as tools for the
detection of expression of the cDNA molecules. Additionally, nucleotide
and protein sequences are useful for suppression of undesired gene
expression in humans and animals. The RdRp is useful as a therapeutic
agent for the control of cancer and virus infection in humans and
animals and the antibody is useful for immunoprecipitation or
immunolocalisation of the protein, identification of polypeptides
interacting with it and screening expression libraries.

Query Match 0.7%; Score 25; DB 23; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.093;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2659 GCCGATGGAATATCTCCAGCACCC 2683
DB 25 GCCGATGGAATATCTCCAGCACCC 1

RESULT 9
AAD04372/c
ID AAD04372 standard; DNA; 25 BP.
XX AC AAD04372;
XX DT 04-JUL-2001 (first entry)
XX DE RACE PCR primer, GSP420 to amplify tomato RdRp cDNA.

```

XX Tomato: gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;  
 KW transgenic plant; tissue culture; plant breeding; therapy; C-protein;  
 KW PCR primer; RACE; ss.  
 XX Lycopersicon esculentum.  
 XX US6218142-B1.  
 XX 17-APR-2001.  
 XX 05-MAR-1997; 97US-0811583.  
 XX 05-MAR-1997; 97US-0811583.  
 XX (WASS/) WASSENEGGER M.  
 XX (RIED/) RIEDEL L.  
 XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;  
 XX WPI; 2001-289830/30.  
 XX New nucleic acid molecules encoding polypeptides with RNA-directed RNA  
 PT polymerase enzymatic activity, useful in modulating gene expression in  
 PT plants, humans and animals, as well as in plant cell/tissue cultures or  
 PT plant breeding -  
 XX Example 3; Column 20; 31pp; English.  
 PS The present sequence is RACE (rapid amplification of 5'-cDNA ends) PCR  
 CC gene specific reverse (GSP) primer, GSP420, used to amplify a cDNA  
 CC encoding tomato C-protein having RNA-directed RNA polymerase (RdRP)  
 CC activity. This protein is capable of RNA-directed RNA synthesis, thus  
 CC using RNA as a template for synthesizing complementary RNA molecules.  
 CC RdRP nucleic acid is useful for modulating gene expression in plants,  
 CC humans and animals. This may lead to various physiological, developmental  
 CC and/or morphological changes. Transgenic plants containing RdRP nucleic  
 CC acid is especially useful in plant cell or tissue cultures and in plant  
 CC breeding. RdRP is useful in gene therapy, particularly for treating a  
 CC disease that is caused by the undesirable expression or overexpression of  
 CC a gene.  
 XX Sequence 25 BP; 9 A; 9 C; 4 G; 3 T; 0 other;  
 SQ  
 Query Match 0.6%; Score 24; DB 22; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.3;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1879 TGTGGGTGTGATCCGGATTTCATC 1902  
 DB 24 TGTGGGTGTGATCCGGATTTCATC 1  
 RESULT 10  
 AAS17839/c  
 ID AAS17839 standard; DNA; 25 BP.  
 XX AAS17839;  
 XX 08-MAY-2002 (first entry)  
 DE Tomato RNA-directed RNA polymerase (RdRP) GSP420 PCR primer.  
 XX Tomato: RdRP; RNA-directed RNA polymerase; in vitro transcription;  
 KW cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;  
 KW transgenic plant; transgenic animal; cancer; viral infection; RACE;  
 KW immunoprecipitation; immunolocalisation; PCR primer; GSP420;  
 KW gene therapy; ss.  
 XX Lycopersicon esculentum.  
 XX US2001023067-A1.

PD 20-SEP-2001.  
 XX 08-FEB-2001; 2001US-0782874.  
 XX 05-MAR-1997; 97US-0811583.  
 XX (WASS/) WASSENEGGER M.  
 XX (RIED/) RIEDEL L.  
 XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;  
 XX WPI; 2001-595798/67.  
 XX New nucleic acid molecule encoding a polypeptide having the enzymatic  
 PT activity of RNA-directed RNA polymerase, for modulating gene expression  
 PT and treating cancer and virus infection in human and animals -  
 XX Example 3; Page 11; 34pp; English.  
 PS This sequence represents the tomato RNA-directed RNA polymerase (RdRP)  
 CC specific PCR primer GSP420 used to obtain the 5' end of the cDNA  
 CC molecule using the rapid amplification of cDNA ends (RACE) technique.  
 CC The invention comprises the nucleic acid and protein sequences of RdRP,  
 CC the protein of the invention can catalyze in vitro transcription of  
 CC short single stranded RNAs into DNA molecules, this transcription can be  
 CC either primed by RNA or DNA oligonucleotides or be unprimed. The protein  
 CC may have cytosolic or virucide activities. The sequences of the  
 CC invention may be used in gene therapy or as an RNA directed RNA  
 CC synthesis inhibitor. The RdRP cDNA sequence and a template nucleic acid  
 CC molecule derived from a nucleic acid molecule which causes a disease are  
 CC useful for treating a disease caused by the undesired expression or  
 CC overexpression of a nucleic acid molecule in a human, rat or mouse, by  
 CC administering the molecules. This system can be used in the preparation  
 CC of a pharmaceutical composition and for inhibiting expression of any  
 CC desired gene by transferring the RdRP system to organisms that either  
 CC lack a comparable mechanism or do not sufficiently express their own  
 CC RdRP. An antibody or an antagonist or inhibitor to the protein are  
 CC useful for inhibiting RNA directed RNA synthesis and for ensuring stable  
 CC heterologous, gene expression in transgenic organisms. The sequence is  
 CC useful for probes and/or for the control of gene expression, as primers  
 CC for amplification of nucleic acid molecules and as tools for the  
 CC detection of expression of the cDNA molecules. Additionally, nucleotide  
 CC and protein sequences are useful for suppression of undesired gene  
 CC expression in humans and animals. The RdRP is useful as a therapeutic  
 CC agent for the control of cancer and virus infection in humans and  
 CC animals and the antibody is useful for immunoprecipitation or  
 CC immunolocalisation of the protein, identification of polypeptides  
 CC interacting with it and screening expression libraries.  
 XX Sequence 25 BP; 9 A; 9 C; 4 G; 3 T; 0 other;  
 SQ  
 Query Match 0.6%; Score 24; DB 23; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.3;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1879 TGTGGGTGTGATCCGGATTTCATC 1902  
 DB 24 TGTGGGTGTGATCCGGATTTCATC 1  
 RESULT 11  
 ABL64204/c  
 ID ABL64204 standard; DNA; 407 BP.  
 XX ABL64204;  
 XX 15-MAY-2002 (first entry)  
 DE Stomach cancer related gene sequence SEQ ID NO:2541.  
 XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
 KW cytosolic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;

gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200194629-A2.  
XX  
PD 13-DEC-2001.  
XX  
PF 30-MAY-2001; 2001WO-US10838.  
XX  
PR 05-JUN-2000; 2000US-209473P.  
PR 05-JUN-2000; 2000US-209531P.  
PR 18-SEP-2000; 2000US-233133P.  
PR 18-SEP-2000; 2000US-233617P.  
PR 20-SEP-2000; 2000US-234009P.  
PR 20-SEP-2000; 2000US-234034P.  
PR 20-SEP-2000; 2000US-234052P.  
PR 22-SEP-2000; 2000US-234509P.  
PR 22-SEP-2000; 2000US-234567P.  
PR 25-SEP-2000; 2000US-234923P.  
PR 25-SEP-2000; 2000US-234924P.  
PR 25-SEP-2000; 2000US-235077P.  
PR 25-SEP-2000; 2000US-235082P.  
PR 25-SEP-2000; 2000US-235134P.  
PR 25-SEP-2000; 2000US-235280P.  
PR 26-SEP-2000; 2000US-235637P.  
PR 26-SEP-2000; 2000US-235638P.  
PR 27-SEP-2000; 2000US-235711P.  
PR 27-SEP-2000; 2000US-235720P.  
PR 27-SEP-2000; 2000US-235840P.  
PR 27-SEP-2000; 2000US-235863P.  
PR 28-SEP-2000; 2000US-236028P.  
PR 28-SEP-2000; 2000US-236032P.  
PR 28-SEP-2000; 2000US-236033P.  
PR 28-SEP-2000; 2000US-236034P.  
PR 28-SEP-2000; 2000US-236109P.  
PR 28-SEP-2000; 2000US-236111P.  
PR 29-SEP-2000; 2000US-236842P.  
PR 29-SEP-2000; 2000US-236891P.  
PR 02-OCT-2000; 2000US-237172P.  
PR 02-OCT-2000; 2000US-237173P.  
PR 02-OCT-2000; 2000US-237278P.  
PR 02-OCT-2000; 2000US-237294P.  
PR 02-OCT-2000; 2000US-237295P.  
PR 03-OCT-2000; 2000US-237316P.  
PR 03-OCT-2000; 2000US-237425P.  
PR 03-OCT-2000; 2000US-237598P.  
PR 03-OCT-2000; 2000US-237604P.  
PR 03-OCT-2000; 2000US-237606P.  
PR 03-OCT-2000; 2000US-237608P.  
PR 01-NOV-2000; 2000US-244867P.  
PR 01-NOV-2000; 2000US-245084P.  
XX  
PA (AVAL-) AVALON PHARM.  
XX  
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX  
XX WPI: 2002-188264/24.  
XX  
XX Screening for anti-neoplastic agent involves exposing cells to a  
PT chemical agent to be tested for anti-neoplastic activity, and  
PT determining a change in expression of a gene of a signature gene set -  
XX  
XX Claim 1; SEQ ID 2541; 44pp; English.  
XX  
XX The present invention describes a method (M1) for screening for an  
CC anti-neoplastic agent. The method involves exposing cells to a chemical  
CC agent to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic

activity and can be used in gene therapy. M1 can be used for screening  
an anti-neoplastic agent, and can be used for producing a product which  
is the data collected with respect to the anti-neoplastic agent as a  
result of M1, and the data is sufficient to convey the chemical  
structure and/or properties of the agent. M1 can be used in the  
treatment of cancer such as colon, breast, stomach, lung, thyroid,  
oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
carcinoma, papillary carcinoma and Wilms' tumour.  
Sequence 407 BP; 120 A; 98 C; 92 G; 97 T; 0 other;  
Query Match 0.5%; Score 20; DB 24; Length 407;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1916 TTGAGAAAGAGCATGTCGAA 1935  
|||||  
Db 384 TTGAGAAAGAGCATGTCGAA 365  
RESULT 12  
ABQ55295  
ID ABQ55295 standard; cDNA; 786 BP.  
XX  
AC ABQ55295;  
XX  
DT 22-AUG-2002 (first entry)  
XX  
DE Human ovarian antigen HCOB20 cDNA, SEQ ID NO:1175.  
XX  
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;  
KW inflammatory condition; immune disorder; blood disorder;  
KW cardiovascular disorder; respiratory disorder; neurological disorder;  
KW gastrointestinal disorder; urinary system disorder; drug screening;  
KW gene therapy; chromosome mapping; forensic analysis;  
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
KW antiinflammatory; gynaecological; reproductive; gene; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200200677-A1.  
PN  
XX 03-JAN-2002.  
PD  
XX 07-JUN-2001; 2001WO-US18569.  
PF  
XX 07-JUN-2000; 2000US-209467P.  
PR  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX Birse CE, Rosen CA;  
PI  
XX WPI: 2002-147878/19.  
DR  
XX P-PSDE; ABP42218.  
DR  
XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
PT useful in the prevention, treatment and diagnosis of cancer (e.g.  
PT ovarian cancer), immune disorders, cardiovascular disorders and  
PT neurological diseases -  
XX  
XX Claim 1; SEQ ID NO 1175; 2922pp; English.  
XX  
XX The invention relates to 2175 novel human ovarian antigens (ABP41054-  
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
CC to the sequences of the invention. The invention additionally relates to  
CC recombinant vectors and host cells comprising human ovarian antigen  
CC polynucleotides, antibodies against human ovarian antigens, and the use

of ovarian antigen polynucleotides and polypeptides in diagnosing, treating, prognosing or preventing various ovary and/or breast-related disorders. Such conditions include ovarian cancer and breast cancer, and metastatic tumours of ovarian or breast origin, reproductive system disorders (e.g., infertility, disorders of pregnancy, anovulation, polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and vaginitis), immune disorders (e.g., congenital and acquired immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus), immunodeficiencies, autoimmune oophoritis, cardiovascular disorders, blood-related disorders (e.g., anaemia), gastrointestinal disorders, respiratory disorders, neurological disorders, gastrointestinal disorders and urinary system disorders. Ovarian antigen polypeptides and polynucleotides may also be used in screening for compounds which modulate ovarian antigen expression or activity. The polynucleotides may further be used for gene therapy, chromosome mapping, in the identification of individuals and in forensic analysis, and the polypeptides may be used as food additives or to prepare antibodies useful in disease diagnosis, drug targeting and phenotyping. The present sequence represents cDNA encoding a human ovarian antigen of the invention.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 786 BP; 184 A; 198 C; 190 G; 198 T; 16 other;

Query Match 0.5%; Score 20; DB 24; Length 786;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1916 TTGAGAAAGAGCATGTCGAA 1935  
|||||  
DB 198 TTGAGAAAGAGCATGTCGAA 217

## RESULT 13

AD38196/c  
ID AAD38196 standard; DNA; 1000 BP.

XX AC AAD38196;

XX DT 10-SEP-2002 (first entry)

XX DE Non-typeable Haemophilus influenzae strain 3224A BASB205 DNA #7.

XX KW BASB205 protein; infection; otitis media; pneumonia; inflammation;  
XX KW invasive disease; hearing loss; fluid accumulation; nerve damage;  
XX KW vaccine; sinusitis; ds.

XX OS Haemophilus influenzae.

XX PN WO200230971-A2.

XX PD 18-APR-2002.

XX PF 05-OCT-2001; 2001WO-EP11560.

XX PR 13-OCT-2000; 2000GB-0025171.

XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Thonnard J;

XX DR WPI; 2002-452340/48.

XX PT Novel isolated non-typeable Haemophilus influenzae BASB205 polypeptides  
XX PT useful as components of vaccines useful for treating bacterial  
XX PT infection such as otitis media, delayed speech learning, inflammation  
XX PT of middle ear -

XX PS Disclosure; Page 80; 92pp; English.

XX

CC The invention relates to non-typeable Haemophilus influenzae BASB205  
CC polypeptides and polynucleotides. Sequences of the invention are used  
CC to treat non-typeable H. influenzae infections. They are also used  
CC as components of vaccines useful for treating infections preferably  
CC bacterial infections such as otitis media, pneumonia, sinusitis,  
CC nosocomial infections, invasive diseases, chronic otitis media with  
CC hearing loss, fluid accumulation in the middle ear, auditive nerve  
CC damage, delayed speech learning, infections of the upper respiratory  
CC tract and inflammation of the middle ear. The present sequence is  
CC non-typeable H. influenzae strain 3224A BASB205 DNA upstream region.

XX SQ Sequence 1000 BP; 288 A; 151 C; 256 G; 305 T; 0 other;

Query Match 0.5%; Score 20; DB 24; Length 1000;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2414 TCCACCAACAGTAATTCAT 2433  
|||||  
DB 879 TCCACCAACAGTAATTCAT 860

## RESULT 14

AAZ52915

ID AAZ52915 standard; cDNA; 1509 BP.

XX AC AAZ52915;

XX DT 14-MAR-2000 (first entry)

XX DE Human prostate tumor cDNA library derived EST fragment #58.

XX KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;  
XX KW treatment; ds.

XX OS Homo sapiens.

XX PN DEL9820190-A1.

XX PD 04-NOV-1999.

XX PF 28-APR-1998; 98DE-1020190.

XX PR 28-APR-1998; 98DE-1020190.

XX PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

XX DR WPI; 1999-621386/54.

XX P-PSDB; AAY73986, AAY73987, AAY73988.

XX PT New human nucleic acid sequences from pancreatic tumors, and related  
XX PT proteins -

XX PS Claim 2; Page 229-230; 502pp; German.

XX This invention describes novel polypeptides and their encoding nucleic  
XX acids derived from human pancreatic tumor tissue which have cytostatic  
XX activity. The sequences are also useful in producing pharmaceutical  
XX compositions for treatment of pancreatic tumors. AAZ52958-253014  
XX represent expressed sequence tag (EST) fragments derived from a human  
XX pancreatic tumor cDNA library and which encode the proteins represented  
XX in AAY73814-Y74252.

XX SQ Sequence 1509 BP; 416 A; 338 C; 416 G; 339 T; 0 other;

Query Match 0.5%; Score 20; DB 20; Length 1509;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1916 TTGAGAAAGAGCATGTCGAA 1935  
|||||



Db 1187 TTGAGAAAGAGCATGTCGAA 1206

RESULT 15

AA18199

ID AAX18199 standard; cDNA to mRNA; 1730 BP.

XX AC AAX18199;

XX DT 10-MAY-1999 (first entry)

XX DE TIP49 coding sequence #2.

XX KW TIP49; TATA binding protein; TBP binding protein;

XX KW DNA transcription control; ss.

XX OS Rattus sp.

XX PN W09900419-A1.

XX PD 07-JAN-1999.

XX PF 25-JUN-1998; 98WO-JP02836.

XX PR 27-JUN-1997; 97JP-0187398.

XX PA (SUME) SUMITOMO ELECTRIC IND CO.

XX PI Kishimoto T, Makino Y, Niwa S, Tamura T;

XX DR WPI; 1999-095682/08.

XX DR P-PSDB; AAW74417.

XX TTBP-binding protein with DNA helicase and ATPase activities - gene

XX PT encoding it, and antibodies recognising it.

XX PS Claim 10; Page 43-45; 64pp; Japanese.

XX CC This sequence encodes the TATA-binding protein (TBP) binding protein,

XX CC designated TIP49, of the invention. TIP49 and its associated nucleic

XX CC acids and antibodies are useful in investigation of the process of DNA

XX CC transcription control by TBP in vivo, and in detection of the blocking of

XX CC DNA transcription.

XX SQ Sequence 1730 BP; 469 A; 401 C; 476 G; 384 T; 0 other;

Query Match 0.5%; Score 20; DB 20; Length 1730;

Best Local Similarity 100.0%; Pred. No. 32;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1916 TTGAGAAAGAGCATGTCGAA 1935

Db 1333 TTGAGAAAGAGCATGTCGAA 1352

RESULT 16

AAC59220

ID AAC59220 standard; cDNA; 1744 BP.

XX AC AAC59220;

XX DT 29-JAN-2001 (first entry)

XX DE Human secreted protein cDNA sequence #6.

XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;

XX KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;

XX KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;

XX KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;

XX KW neurological disease; infection; human; secreted protein; ss.

XX OS Homo sapiens.

XX

PN W0200055199-A1.

XX 21-SEP-2000.

XX 09-MAR-2000; 2000WO-US06014.

XX 12-MAR-1999; 99US-0124095.

XX PR 11-JUN-1999; 99US-0138598.

XX PR 03-DEC-1999; 99US-0168665.

XX (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM, Komatsoulis G;

XX DR WPI; 2000-572359/53.

XX DR P-PSDB; AAB27799.

XX Isolated nucleic acid molecule encoding a human secreted protein is

XX used in preventing, treating or ameliorating a medical condition -

XX Claim 1; Page 343; 433pp; English.

XX The invention relate to the isolation of genes AAC59215-C59261 encoding

XX 47 human secreted proteins AAB27794-B27840. The genes can be used to

XX generate fusion proteins by linking to the gene for the human

XX immunoglobulin G Fc portion for increasing the stability of

XX the fusion protein as compared to the human protein only. The genes and

XX proteins are useful for preventing, ameliorating or treating medical

XX conditions, e.g. by protein or gene therapy. The genes are isolated

XX from a range of human tissues disclosed in the specification. The

XX nucleic acids, proteins, antibodies and (ant)agonists are useful in

XX the diagnosis, treatment and prevention of: (a) cancer, e.g. breast

XX and ovarian cancer, and other cancers of the adrenal gland, bone, bone

XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;

XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune

XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's

XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative

XX colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)

XX wound healing; (e) neurological diseases e.g. cerebral anoxia and

XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal

XX and parasitic infections.

XX SQ Sequence 1744 BP; 476 A; 403 C; 480 G; 384 T; 1 other;

Query Match 0.5%; Score 20; DB 21; Length 1744;

Best Local Similarity 100.0%; Pred. No. 32;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1916 TTGAGAAAGAGCATGTCGAA 1935

Db 1342 TTGAGAAAGAGCATGTCGAA 1361

RESULT 17

AAX24059

ID AAX24059 standard; DNA; 1750 BP.

XX AC AAX24059;

XX DT 30-JUN-1999 (first entry)

XX DE Human regulatory molecule HRM-1 DNA.

XX KW Human regulatory molecule; HRM-1; cytostatic activity; immune modulator;

XX KW transcription factor; enhancer; cell proliferation stimulation; cancer;

XX KW treatment; microarray; detection; diagnosis; cell proliferation disease;

XX KW leukemia; lymphoma; myeloma; adenocarcinoma; sarcoma; bladder; bone;

XX KW brain; lung; liver; ovary; skin; teratocarcinoma; immune response;

XX KW allergy; asthma; diabetes; multiple sclerosis; Grave's disease;

XX KW myasthenia gravis; ss.

XX OS Homo sapiens.

XX

```

PN WO9915658-A2.
PD
XX 01-APR-1999.
XX
XX 22-SEP-1998; 98WO-US19839.
XX
XX 23-SEP-1997; 97US-0933750.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Au-Young J, Bandman O, Guegler KJ, Hillman JL, Lal P;
PI Shah P, Yue H;
XX
XX WPI; 1999-254710/21.
DR P-PSDB; AAW93945.
XX
XX New human regulatory molecules
PT
XX
XX Claim 5; Page 70-71; 76pp; English.
XX
XX This invention describes novel human regulatory molecules (HRM) which
CC have cytostatic activity and act as immune modulators, transcription
CC factors or enhancers. The HRMs can be used to stimulate cell
CC proliferation. Antagonists and agonists of the proteins of the invention
CC can be used to treat cancer. The encoding nucleic acids can be used in
CC microarrays to detect polynucleotides (and their expression levels) that
CC encode HRMs in a biological sample. The HRMs and microarrays can be used
CC to diagnose, treat or prevent cell proliferation diseases especially cancer,
CC e.g. leukemia, lymphoma, myeloma, adenocarcinoma, sarcoma, cancer of e.g.
CC bladder, bone, brain, lung, liver, ovary, skin, etc, teratocarcinoma, or
CC to treat or prevent immune responses e.g. allergies, asthma, diabetes,
CC multiple sclerosis, Grave's disease or myasthenia gravis.
XX
XX Sequence 1750 BP; 472 A; 406 C; 485 G; 387 T; 0 other;
SQ
Query Match 0.5%; Score 20; DB 20; Length 1750;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1916 TTGAGAAAGAGCATGTCGAA 1935
Db 1355 TTGAGAAAGAGCATGTCGAA 1374
|||||
RESULT 18
ABA92477
ID ABA92477 standard; cDNA; 1750 BP.
XX
XX ABA92477;
XX
XX 18-MAR-2002 (first entry)
XX
XX Human TIP49 encoding cDNA.
XX
XX Human; STAP1; Skp2; cyclin A-interacting protein; TIP48; TIP49;
KW Skp2-associated protein one; anti-cancer; anti-proliferative; cancer;
KW cytostatic; transcription regulation; non-cMyc expressing cancer; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 77..1447
XX /*tag= a
XX /*product= "TIP49"
XX
XX WO200185980-A2.
XX
XX 15-NOV-2001.
XX
XX 11-MAY-2001; 2001WO-EP05404.
XX
XX 12-MAY-2000; 2000GB-0011439.
XX
XX
XX (NOVS ) NOVARTIS FORSCHUNGSSTIFTUNG ZWEIGNIEDERL.
XX
XX Gstaiger MGC, Krek W;
XX
XX WPI; 2002-049448/06.
XX
XX P-PSDB; ABB04982.
XX
XX Complex containing a TIP48 or a TIP49 family member and various other
XX proteins is useful to screen for new agents to treat cancers that do
XX not express cMyc
XX
XX Example 8; Fig 4; 34pp; English.
XX
XX The present invention describes a complex (I) comprising a TIP49 family
XX member and STAP1 (Skp1(a cyclin A-interacting protein)-associated
XX protein one), prefoldin, RPB5 and/or RPL1, preferably in a 1:1:1:1
XX ratio. Also described are: (1) a transcription regulatory complex
XX comprising TIP48 and/or TIP49 and three or more proteins or
XX polypeptides; (2) identifying an agent active against cancer cells that
XX do not express c-Myc, comprising contacting a member of the TIP49 family
XX with a test compound and measuring TIP49 enzymatic or ligand binding
XX activity; and (3) an anti-cancer agent, preferably an anti-proliferative
XX agent, particularly a nucleic acid or antibody identified by the method
XX of (2). (1) has cytostatic activity, and can be used in transcription
XX regulation. (I) can be used for screening new agents for the treatment
XX of non-cMyc expressing cancers. The present sequence encodes human
XX TIP49 from the present invention.
XX
XX Sequence 1750 BP; 476 A; 403 C; 484 G; 387 T; 0 other;
SQ
Query Match 0.5%; Score 20; DB 24; Length 1750;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1916 TTGAGAAAGAGCATGTCGAA 1935
Db 1350 TTGAGAAAGAGCATGTCGAA 1369
|||||
RESULT 19
AAK72348
ID AAK72348 standard; DNA; 38653 BP.
XX
XX AAK72348;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27160.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
XX
XX 24-FEB-2000; 2000US-0184664.
XX
XX 02-MAR-2000; 2000US-0186350.
XX
XX 16-MAR-2000; 2000US-0189874.
XX
XX 17-MAR-2000; 2000US-0190076.
XX
XX 18-APR-2000; 2000US-0198123.
XX
XX 19-MAY-2000; 2000US-0205515.
XX
XX 07-JUN-2000; 2000US-0209467.
XX
XX 28-JUN-2000; 2000US-0214886.
XX
XX 30-JUN-2000; 2000US-0215135.
XX
XX 07-JUL-2000; 2000US-0216647.
XX
XX 07-JUL-2000; 2000US-0216880.

```

PR 11-JUL-2000; 2000US-02117487.  
PR 11-JUL-2000; 2000US-02117496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0225279.  
PR 22-AUG-2000; 2000US-0225681.  
PR 22-AUG-2000; 2000US-0225688.  
PR 22-AUG-2000; 2000US-0225689.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233065.  
PR 14-SEP-2000; 2000US-0233066.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235634.  
PR 27-SEP-2000; 2000US-0235636.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251889.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-483426/52.  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating cancers and  
XX metastasis -  
XX Disclosure; SEQ ID NO 27160; 3071pp + Sequence Listing; English.  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
XX activity, and can be used in gene therapy and vaccine production. (I)  
XX proteins, and polynucleotides may be used in the prevention, diagnosis and  
XX treatment of diseases associated with inappropriate (I) expression. For  
XX example, they may be used to treat disorders associated with decreased  
XX expression by rectifying mutations or deletions in a patient's genome  
XX that affect the activity of (I) by expressing inactive proteins or to  
XX supplement the patients own production of (I). Additionally, (I)  
XX polynucleotides may be used to produce the secreted (I), by inserting

CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (1) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK6794 represent human immune/haematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX

SQ Sequence 38653 BP; 10472 A; 7264 C; 8237 G; 12680 T; 0 other;

Query Match 0.5%; Score 20; DB 22; Length 38653;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2612 TACTTGTGTTGGGATCA 2631  
 |||||  
 DB 11586 TACTTGTGTTGGGATCA 11605

RESULT 20  
 AAD04374  
 ID AAD04374 standard; DNA; 34 BP.  
 XX  
 AC AAD04374;  
 XX  
 DT 04-JUL-2001 (first entry)  
 XX  
 DE Forward PCR primer, P127Bam1 to amplify tomato RdRP cDNA.

XX Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;  
 KW transgenic plant; tissue culture; plant breeding; therapy; C-protein;  
 KW PCR primer; ss.  
 XX

OS Lycopersicon esculentum.

XX

PN US6218142-B1.

XX

PD 17-APR-2001.

XX

PF 05-MAR-1997; 97US-0811583.

XX

PR 05-MAR-1997; 97US-0811583.

XX

PA (WASS/) WASSENEGGER M.

XX

PI (RIED/) RIEDEL L.

XX

PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;

XX

DR WPI; 2001-289830/30.

XX

PT New nucleic acid molecules encoding polypeptides with RNA-directed RNA

XX

PT polymerase enzymatic activity, useful in modulating gene expression in

XX

PT plants, humans and animals, as well as in plant cell/tissue cultures or

XX

PT plant breeding -

XX  
 PS Example 3; Column 21; 31pp; English.  
 XX  
 CC The present sequence is forward PCR primer, P127Bam1, used to amplify a  
 CC cDNA encoding tomato C-protein having RNA-directed RNA polymerase (RdRP)  
 CC activity. This primer contains BamHI site to enhance the cloning  
 CC efficiency. This protein is capable of RNA-directed RNA synthesis, thus  
 CC using RNA as a template for synthesising complementary RNA molecules.  
 CC RdRP nucleic acid is useful for modulating gene expression in plants,  
 CC humans and animals. This may lead to various physiological, developmental  
 CC and/or morphological changes. Transgenic plants containing RdRP nucleic  
 CC acid is especially useful in plant cell or tissue cultures and in plant  
 CC breeding. RdRP is useful in gene therapy, particularly for treating a  
 CC disease that is caused by the undesirable expression or overexpression of  
 CC a gene.  
 XX  
 SQ Sequence 34 BP; 8 A; 15 C; 4 G; 7 T; 0 other;

Query Match 0.5%; Score 19; DB 22; Length 34;

Best Local Similarity 100.0%; Pred. No. 95;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 ACTATCACTCCCTCAAG 51  
 |||||  
 DB 16 ACTATCACTCCCTCAAG 34

RESULT 21

AAS17841

ID AAS17841 standard; DNA; 34 BP.

XX

AC AAS17841;

XX

DT 08-MAY-2002 (first entry)

XX

DE Tomato RNA-directed RNA polymerase (RdRP) P127Bam1 PCR primer.

XX

KW Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;

XX

KW cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;

XX

KW transgenic plant; transgenic animal; cancer; viral infection; RACE;

XX

KW immunoprecipitation; immunolocalisation; PCR primer; P127Bam1;

XX

OS Lycopersicon esculentum.

XX

PN US2001023067-A1.

XX

PD 20-SEP-2001.

XX

PF 08-FEB-2001; 2001US-0782874.

XX

PR 05-MAR-1997; 97US-0811583.

XX

PA (WASS/) WASSENEGGER M.

XX

PI (RIED/) RIEDEL L.

XX

PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;

XX

DR WPI; 2001-595798/67.

XX

PT New nucleic acid molecule encoding a polypeptide having the enzymatic

XX

PT activity of RNA-directed RNA polymerase, for modulating gene expression

XX

and treating cancer and virus infection in human and animals -

XX  
 PS Example 3; Page 12; 34pp; English.

XX

CC This sequence represents the tomato RNA-directed RNA polymerase (RdRP)

CC specific PCR primer P127Bam1 used to clone the full length RdRP cDNA

CC molecule of the invention. The invention comprises the nucleic acid and

CC protein sequences of RdRP, the protein of the invention can catalyse in

CC vitro transcription of short single stranded RNAs into DNA molecules.

CC this transcription can be either primed by RNA or DNA oligonucleotides

CC or be unprimed. The protein may have cytosolic or virucide activities.

CC The sequences of the invention may be used in gene therapy or as an RNA

CC directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template

CC nucleic acid molecule derived from a nucleic acid molecule which causes

CC a disease are useful for treating a disease caused by the undesired

CC expression or overexpression of a nucleic acid molecule in a human, rat

CC or mouse, by administering the molecules. This system can be used in the

CC preparation of a pharmaceutical composition and for inhibiting the

CC expression of any desired gene by transferring the RdRP system to

CC organisms that either lack a comparable mechanism or do not sufficiently

CC express their own RdRP. An antibody or an antagonist or inhibitor to the

CC protein are useful for inhibiting RNA directed RNA synthesis and for

CC ensuring stable heterologous gene expression in transgenic organisms.

CC The sequence is useful for probes and/or for the control of gene

CC expression, as primers for amplification of nucleic acid molecules and

CC as tools for the detection of expression of cDNA molecules.

CC Additionally, nucleotide and protein sequences are useful for

CC suppression of undesired gene expression in humans and animals. The RdRP

CC is useful as a therapeutic agent for the control of cancer and virus

CC infection in humans and animals and the antibody is useful for

CC immunoprecipitation or immunolocalisation of the protein, identification  
CC of polypeptides interacting with it and screening expression libraries.

XX SQ Sequence 34 BP; 8 A; 15 C; 4 G; 7 T; 0 other;

Query Match 0.5%; Score 19; DB 23; Length 34;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 ACTCATCTACCTCCCTCAAG 51  
|||||  
Db 16 ACTCATCTACCTCCCTCAAG 34

RESULT 22  
AAH81903/c  
ID AAH81903 standard; DNA; 255 BP.

XX AC AAH81903;

XX DT 21-SEP-2001 (first entry)

XX DE Rat differential transcription-associated cDNA SEQ ID 412.

XX KW Differential transcription; human; rat; tumour cell; cytostatic;  
XX KW Ras modulator; Class II tumour suppressor gene; gene therapy; ss.

XX OS Rattus sp.

XX PN WO200157058-A2.

XX PD 09-AUG-2001.

XX PF 31-JAN-2001; 2001WO-EP01003.

XX PR 31-JAN-2000; 2000DE-1004102.

XX PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX PI Rosenthal A, Hinzmann B, Schaefer R, Zuber J, Tchernitsa O;

XX PI Grips M, Hellriegel M, Schmitz A, Sers C;

XX DR WPI; 2001-483415/52.

XX Nucleic acids differentially expressed between tumor and normal cells,  
PT useful for diagnosis or therapy of tumors and for screening active  
PT agents -

XX Claim 6; Page 492; 579pp; German.

XX CC This invention describes a nucleic acid (I) with differential expression  
CC between tumor and normal cells and which has cytostatic activity. (I)  
CC work as modulators of Ras activity by inducing expression of tumour  
CC suppressor genes. (I), and polypeptides encoded by them, are useful as  
CC targets for diagnosis or therapy and in screening to determine the  
CC effects of an active compound (potential pharmaceutical) on a cell line,  
CC particularly for diagnosis and treatment of tumors, especially by  
CC modulating expression of (I) (by gene therapy, antisense RNA or ribozyme  
CC methods) or by modulating the amount and/or location of (I)-encoded  
CC polypeptides (by administration of the polypeptide or its activator,  
CC antibody (optionally as a conjugate) or inhibitor). The method allows  
CC identification of many class II tumour suppressor genes (i.e. genes that  
CC are not primary targets for tumour-initiating mutations).  
CC AAH81492-AAH82376 represent the human and rat derived nucleic acid  
CC fragments described in the method of the invention.

XX SQ Sequence 255 BP; 98 A; 47 C; 59 G; 51 T; 0 other;

Query Match 0.5%; Score 19; DB 22; Length 255;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1758 ATGCTTTTCTGATGAAT 1776

Db 113 ATGCTTTTCTGATGAAT 95  
|||||

RESULT 23  
ABN67264/c  
ID ABN67264 standard; DNA; 327 BP.

XX AC ABN67264;

XX DT 01-JUL-2002 (first entry)

XX DE Streptococcus polynucleotide SEQ ID NO 2441.

XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.

XX OS Streptococcus agalactiae.

XX PN WO200234771-A2.

XX PD 02-MAY-2002.

XX PF 29-OCT-2001; 2001WO-GB04789.

XX PR 27-OCT-2000; 2000GB-0026333.

XX PR 24-NOV-2000; 2000GB-0028727.

XX PR 07-MAR-2001; 2001GB-0005640.

XX PA (CHIR-) CHIRON SPA.

XX PA (GENO-) INST GENOMIC RES.

XX PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

XX PI Tettelin H;

XX DR WPI; 2002-352536/38.

XX DR P-PSDB; ABP26633.

XX Claim 7; Page 3395; 4525pp; English.

XX CC The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins.

XX SQ Sequence 327 BP; 99 A; 57 C; 74 G; 97 T; 0 other;

Query Match 0.5%; Score 19; DB 24; Length 327;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2416 CACCAACAGTAATTCATT 2434  
|||||

Db 292 CACCAACAGTAATTCATT 274

RESULT 24  
AAC31397/C  
ID AAC31397 standard; cDNA; 369 BP.  
XX AC AAC31397;  
XX DT 06-OCT-2000 (first entry)  
XX DE Human secreted protein 5' EST, SEQ ID NO: 35472.  
XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX KW gene therapy; chromosome mapping; ss.  
XX OS Homo sapiens.  
XX PN EP1033401-A2.  
XX PD 06-SEP-2000.  
XX PF 21-FEB-2000; 2000EP-0200610.  
XX PR 26-FEB-1999; 99US-0122487.  
XX PA (GEST ) GENSET.  
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;  
XX DR WPI; 2000-500381/45.  
XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX PS Claim 1; SEQ ID 35472; 71pp + CD-ROM; English.  
XX CC The present sequence is one of a large number of 5' ESTs derived from  
XX CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
XX CC identified within the present sequence. The 5' ESTs were prepared from  
XX CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
XX CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
XX CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
XX CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
XX CC derived from the 5' ends of mRNAs and even in those cases where longer  
XX CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
XX CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
XX CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
XX CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
XX CC They are used to obtain upstream regulatory sequences and to design  
XX CC expression and secretion vectors.  
XX SQ Sequence 369 BP; 72 A; 76 C; 83 G; 138 T; 0 other;  
Query Match 0.5%; Score 19; DB 21; Length 369;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2065 AAAGGAAGCTGTAGATCAG 2083  
Db 62 AAAGGAAGCTGTAGATCAG 44  
RESULT 25  
AAI87435  
ID AAI87435 standard; cDNA; 452 BP.  
XX AC AAI87435;  
XX DT 06-NOV-2001 (first entry)  
XX DE Human polynucleotide SEQ ID NO 7495.  
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation; ss.  
XX OS Homo sapiens.  
XX PN WO200164835-A2.  
XX PD 07-SEP-2001.  
XX PF 26-FEB-2001; 2001WO-US04927.  
XX PR 28-FEB-2000; 2000US-0515126.  
XX PR 18-MAY-2000; 2000US-0577409.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Tang YT, Liu C, Drmanac RT;  
XX DR WPI; 2001-514838/56.  
XX DR P-PSDB; AAO07504.  
XX PT Isolated nucleic acids and polypeptides, useful for preventing  
XX PT diagnosing and treating e.g. leukaemia, inflammation and immune  
XX PT disorders -  
XX PS Claim 1; SEQ ID NO 7495; 1399pp + Sequence Listing; English.  
XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
XX CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
XX CC cytokine, cell proliferation or cell differentiation or which may induce  
XX CC production of other cytokines in other cell populations. The  
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX CC peptide therapy. The polypeptides have various cytokine-like activities,  
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating  
XX CC activity, tissue growth factor activity, immunomodulatory activity and  
XX CC activin/inhibin activity and may be useful in the diagnosis and/or  
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
XX CC inflammation.  
XX CC Note: The sequence data for this patent did not form part of the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 452 BP; 213 A; 53 C; 77 G; 104 T; 5 other;  
Query Match 0.5%; Score 19; DB 22; Length 452;  
Best Local Similarity 100.0%; Pred. No. 1e-02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 578 AAATTTTCTTTCTTTTAT 596  
Db 390 AAATTTTCTTTCTTTTAT 408  
RESULT 26  
AAI49104/C  
ID AAI49104 standard; DNA; 456 BP.  
XX AC AAI49104;  
XX DT 23-SEP-1997 (first entry)  
XX DE Partial DNA clone AciI#1-426 encoding immunostimulatory peptide.  
XX KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;  
XX KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;  
XX KW tuberculin skin test; ds.  
XX OS Mycobacterium tuberculosis.  
XX PN WO9700067-A1.  
XX PD 03-JAN-1997.

```

XX PF 14-JUN-1996; 95WO-US10375.
XX PF
XX PR 15-JUN-1995; 95US-0000254.
XX PR
XX PA (UYVI-) UNIV VICTORIA.
XX PI
XX PI Nano FE;
XX DR WPI; 1997-077347/07.
XX DR
XX XX
XX PT New immuno-stimulatory peptide(s) of Mycobacterium tuberculosis -
XX PT useful in vaccines, diagnostic skin test, immunoassay and gene
XX PT isolation
XX PS
XX PS Claim 1; Page 34; 79pp; English.
XX CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
XX CC which encode partial sequences of immunostimulatory peptides. Each of
XX CC the clones encode at least one immunostimulatory T cell epitope. The
XX CC clones were identified by testing over 300 fusion clones (alkaline
XX CC phosphatase-M. tuberculosis peptide fusions) for their ability to
XX CC stimulate interferon (IFN)-gamma production. 80 clones were initially
XX CC designated to have some ability to stimulate IFN-gamma production, of
XX CC which 76 are shown in AAT49100-175. These sequences can be used to
XX CC obtain the full length M. tuberculosis genes and corresponding proteins
XX CC using standard techniques. The peptides are useful in vaccines, as
XX CC reagents in an improved tuberculin skin test (especially using peptides
XX CC different from those used in vaccines so as to allow differentiation
XX CC between vaccinated and infected subjects) and as immunoassay reagents
XX CC for detecting specific antibodies. An advantage of these peptides is
XX CC that they stimulate production of IFN-gamma (critical for a protective
XX CC immune response to M. tuberculosis) by CD4-positive T cells. The protein
XX CC encoded by this sequence has amino acid similarity to a dipeptide
XX CC transport protein.
XX
XX SQ Sequence 456 BP; 63 A; 121 C; 165 G; 98 T; 9 other;
      Query Match 0.5%; Score 19; DB 18; Length 456;
      Best Local Similarity 100.0%; Pred. No. 1e+02;
      Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2406 ACAGCAGATCCACCAACAG 2424
      |||||
DB 296 ACAGCAGATCCACCAACAG 278

RESULT 27
ABA59152/c
ID ABA59152 standard; DNA; 461 BP.
XX AC
XX AC ABA59152;
XX DT
XX DT 01-FEB-2002 (first entry)
XX DE
XX DE Human foetal liver single exon nucleic acid probe #7457.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX OS Homo sapiens.
XX XX WO200157277-A2.
XX PN
XX PD
XX PD 09-AUG-2001.
XX PF
XX PF 30-JAN-2001; 2001WO-US00569.
XX XX
XX XX 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483446/52.
XX XX
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains -
XX

```

```

PR 04-OCT-2000; 2000GB-0024263.
XX XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX
XX XX WPI; 2001-483447/52.
XX DR
XX DR Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human fetal liver -
XX PT
XX PS Claim 1; SEQ ID NO 7457; 639pp + sequence listing; English.
XX XX
XX CC The invention relates to a single exon nucleic acid probe for
XX CC measuring human gene expression in a sample derived from human foetal
XX CC liver. The single exon nucleic acid probes may be used for predicting,
XX CC measuring and displaying gene expression in samples derived from human
XX CC fetal liver. The present sequence is a single exon nucleic acid
XX CC probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX XX
XX SQ Sequence 461 BP; 162 A; 78 C; 101 G; 120 T; 0 other;
      Query Match 0.5%; Score 19; DB 22; Length 461;
      Best Local Similarity 100.0%; Pred. No. 1e+02;
      Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 629 TCATATGAAATATATGCG 647
      |||||
DB 56 TCATATGAAATATATGCG 38

RESULT 28
AAK07338/c
ID AAK07338 standard; DNA; 461 BP.
XX XX
XX AC AAK07338;
XX DT
XX DT 05-NOV-2001 (first entry)
XX DE
XX DE Human brain expressed single exon probe SEQ ID NO: 7329.
XX KW Human; brain expressed exon; gene expression analysis; probe;
XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX KW epilepsy; cancer; ss.
XX OS Homo sapiens.
XX XX WO200157275-A2.
XX PN
XX PD
XX PD 09-AUG-2001.
XX PF
XX PF 30-JAN-2001; 2001WO-US00667.
XX XX
XX XX 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483446/52.
XX XX
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains -
XX

```

PS Example 4; SEQ ID NO: 7329; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.

XX Sequence 461 BP; 162 A; 78 C; 101 G; 120 T; 0 other;

Query Match 0.5%; Score 19; DB 22; Length 461;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 TCATATGAAATATATGCG 647  
DB 56 TCATATGAAATATATGCG 38

RESULT 29

AAK33106/c

ID AAK33106 standard; DNA; 461 BP.

XX AC AAK33106;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed single exon probe SEQ ID NO: 7663.

XX Human; bone marrow expressed exon; gene expression analysis; probe;

KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX OS Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00668.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human bone marrow

PS Example 4; SEQ ID NO: 7663; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
CC the probes of the invention.

XX Sequence 461 BP; 162 A; 78 C; 101 G; 120 T; 0 other;

Query Match 0.5%; Score 19; DB 22; Length 461;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 TCATATGAAATATATGCG 647  
DB 56 TCATATGAAATATATGCG 38

RESULT 30

AAI38901/c

ID AAI38901 standard; DNA; 461 BP.

XX AC AAI38901;

XX 17-OCT-2001 (first entry)

XX Probe #7587 used to measure gene expression in human placenta sample.

KW Probe; microarray; human; placenta; antenatal diagnosis;  
genetic disorder; ss.

XX OS Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488997/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta

PS Claim 25; SEQ ID No 7587; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs).  
CC The present sequence is one such probe. The probes are useful for  
CC producing a microarray for predicting, measuring and displaying gene  
CC expression in samples derived from human placenta. The probes are useful  
CC for antenatal diagnosis of human genetic disorders.

XX Sequence 461 BP; 162 A; 78 C; 101 G; 120 T; 0 other;

Query Match 0.5%; Score 19; DB 22; Length 461;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 TCATATGAAATATATGCG 647  
DB 56 TCATATGAAATATATGCG 38

RESULT 31

ABS07923/c

ID ABS07923 standard; DNA; 461 BP.

XX AC ABS07923;

XX 19-AUG-2002 (first entry)

XX Human genome-derived single exon probe from lung SEQ ID No 7914.



Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;  
chronic obstructive pulmonary disease; interstitial lung disease;  
familial idiopathic pulmonary fibrosis; neurofibromatosis;  
tuberoscleriosis; Gaucher's disease; Niemann-Pick disease;  
Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
primary ciliary dyskinesia; pulmonary hypertension;  
hyaline membrane disease.  
Homo sapiens.  
W02001186003-A2.  
15-NOV-2001.  
30-JAN-2001; 2001WO-US00665.  
04-FEB-2000; 2000US-180312P.  
26-MAY-2000; 2000US-207456P.  
30-JUN-2000; 2000US-0608408.  
03-AUG-2000; 2000US-0632366.  
21-SEP-2000; 2000US-234687P.  
27-SEP-2000; 2000US-236359P.  
04-OCT-2000; 2000GB-0024263.  
(MOLE-) MOLECULAR DYNAMICS INC.  
Penn SG, Hanzel DK, Chen W, Rank DR;  
WPI; 2002-114183/15.  
Spatially-addressable set of single exon nucleic acid probes, used to  
measure gene expression in human lung samples -  
Claim 1; SEQ ID No 7914; 634pp; English.  
The invention relates to a spatially-addressable set of single exon  
nucleic acid probes for measuring gene expression in a sample derived  
from human lung comprising single exon nucleic acid probes having one of  
12614 nucleic acid sequences mentioned in the specification, or their  
complements or the 12387 open reading frames derived from the 12614  
probes. Also included are a microarray comprising the novel set of  
probes; the novel set of probes which hybridise at high stringency to a  
nucleic acid expressed in the human lung; measuring gene expression in a  
sample derived from human lung, comprising (a) contacting the array with  
a collection of detectably labeled nucleic acids derived from human lung  
mRNA, and (b) measuring the label detectably bound to each probe of  
the array; identifying exons in a eukaryotic genome, comprising  
(a) algorithmically predicting at least one exon from genomic sequences  
of the eukaryote; and (b) detecting specific hybridisation of detectably  
labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
having a fragment identical to the predicted exon, the probe is included  
in the above mentioned microarray; assigning exons to a single gene,  
comprising (a) identifying exons from genomic sequence by the method  
above and (b) measuring the expression of each of the exons in several  
tissues and/or cell types using hybridisation to a single exon  
microarrays having a probe with the exon, where a common pattern of  
expression of the exons in the tissues and/or cell types indicates that  
the exons should be assigned to a single gene; a peptide comprising one  
of 12011 sequences, mentioned in the specification, or encoded by the  
probes/open reading frames (ORF). The probes are used for gene  
expression analysis, and for identifying exons in a gene, particularly  
using human lung derived mRNA and for the study of lung diseases  
such as asthma, lung cancer, chronic obstructive pulmonary disease  
(COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
and hyaline membrane disease. The present sequence is a single exon  
probe of the invention.

CC Note: The sequence data for this patent did not form part  
of the printed specification, but was obtained in electronic  
format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 461 BP; 162 A; 78 C; 101 G; 120 T; 0 other;  
Query Match 0.5%; Score 19; DB 24; Length 461;  
Best Local Similarity 100.0%; Pred. No. 1e-02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 629 TCATATGAAATATATATGCG 647  
DB 56 TCATATGAAATATATATGCG 38  
RESULT 32  
AAF22601/c  
ID AAF22601 standard; cDNA: 757 BP.  
XX  
AC AAF22601;  
XX  
DT 26-MAR-2001 (first entry)  
XX  
DE Human breast cancer associated antigen nucleotide sequence SEQ ID NO:180.  
XX  
KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;  
KW cancer associated antigen; cytostatic; cancer vaccine; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200073801-A2.  
XX  
PD 07-DEC-2000.  
XX  
PF 26-MAY-2000; 2000WO-US14749.  
XX  
PR 28-MAY-1999; 99US-0136526.  
PR 10-SEP-1999; 99US-0153454.  
XX  
PA (LUDW-) LUDWIG INST CANCER RES.  
XX  
PI Obata Y;  
XX  
DR WPI; 2001-025274/03.  
XX  
PT Nucleic acids encoding breast, gastric and prostate cancer associated  
PT antigen precursors, useful for diagnosing and treating a condition  
PT characterized by expression of an abnormal amount of a protein, e.g.  
PT cancer -  
PS Claim 50; Page 332; 799pp; English.  
XX  
CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014  
CC represent nucleotide sequences encoding human breast, gastric and  
CC prostate cancer associated antigen precursors (CAAP) respectively.  
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970  
CC represent human breast, gastric and prostate CAAP protein sequence  
CC respectively. CAAPs have cytostatic activity and can be used in the  
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic  
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a  
CC condition characterised by expression of an abnormal amount of a protein,  
CC e.g. cancer.  
XX  
SQ Sequence 757 BP; 146 A; 232 C; 244 G; 132 T; 3 other;  
Query Match 0.5%; Score 19; DB 22; Length 757;  
Best Local Similarity 100.0%; Pred. No. 1e-02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2775 ATGCCCATGTCGTATTGC 2793  
DB 257 ATGCCCATGTCGTATTGC 239

RESULT 33  
AAC38539/C  
ID AAC38539 standard; DNA; 783 BP.  
XX AC AAC38539;  
XX AC  
XX AC  
DT 17-OCT-2000 (first entry)  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 21334.  
XX Arabidopsis thaliana.  
XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX Arabidopsis thaliana.  
XX Arabidopsis thaliana.  
XX EP1033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
XX 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125788.  
XX 25-MAR-1999; 99US-0126264.  
XX 29-MAR-1999; 99US-0126785.  
XX 01-APR-1999; 99US-0127462.  
XX 06-APR-1999; 99US-0128234.  
XX 08-APR-1999; 99US-0128714.  
XX 16-APR-1999; 99US-0129845.  
XX 19-APR-1999; 99US-0130077.  
XX 21-APR-1999; 99US-0130449.  
XX 23-APR-1999; 99US-0130510.  
XX 23-APR-1999; 99US-0130891.  
XX 28-APR-1999; 99US-0131449.  
XX 30-APR-1999; 99US-0132048.  
XX 30-APR-1999; 99US-0132407.  
XX 04-MAY-1999; 99US-0132484.  
XX 05-MAY-1999; 99US-0132485.  
XX 06-MAY-1999; 99US-0132486.  
XX 07-MAY-1999; 99US-0132487.  
XX 07-MAY-1999; 99US-0132863.  
XX 11-MAY-1999; 99US-0134256.  
XX 14-MAY-1999; 99US-0134218.  
XX 14-MAY-1999; 99US-0134219.  
XX 14-MAY-1999; 99US-0134321.  
XX 14-MAY-1999; 99US-0134370.  
XX 18-MAY-1999; 99US-0134768.  
XX 19-MAY-1999; 99US-0134941.  
XX 20-MAY-1999; 99US-0135124.  
XX 21-MAY-1999; 99US-0135353.  
XX 24-MAY-1999; 99US-0135629.  
XX 25-MAY-1999; 99US-0136021.  
XX 27-MAY-1999; 99US-0136592.  
XX 28-MAY-1999; 99US-0136782.  
XX 01-JUN-1999; 99US-0137222.  
XX 03-JUN-1999; 99US-0137528.  
XX 04-JUN-1999; 99US-0137502.  
XX 07-JUN-1999; 99US-0137724.  
XX 08-JUN-1999; 99US-0138094.  
XX 10-JUN-1999; 99US-0138540.  
XX 10-JUN-1999; 99US-0138847.  
XX 14-JUN-1999; 99US-0139119.  
XX 16-JUN-1999; 99US-0139452.  
XX 16-JUN-1999; 99US-0139453.  
XX 17-JUN-1999; 99US-0139492.  
XX 18-JUN-1999; 99US-0139454.  
XX 18-JUN-1999; 99US-0139455.  
XX 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.

```
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.5%; Score 19; DB 21; Length 783;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2094 TCTGTCATGATCTTTGAA 2112
DB 597 TCTGTCATGATCTTTGAA 579

RESULT 34
AAI94184
ID AAI94184 standard; cDNA; 889 BP.
XX
AC AAI94184;

us-09-782-874-1.oli.rng

XX 13-NOV-2001 (first entry)
XX Human neuroblastoma expressed polynucleotide SEQ ID NO 259.
XX Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX Homo sapiens.
XX WC200166719-A1.
XX 13-SEP-2001.
XX 02-MAR-2001; 2001WO-JP01629.
XX 07-MAR-2000; 2000JP-0159195.
XX (CHIB-) CHIBA PREFECTURE.
XX (HISM) HISAMITSU PHARM CO LTD.
XX Nakagawara A;
XX WPI; 2001-565584/63.
XX Nucleic acids originating in gene expressed in human neuroblastoma,
XX useful as probe or primer in diagnosing prognosis of human
XX neuroblastoma, malignancy and susceptibility indicator or tumour marker
XX for anti-cancer agents.
XX Claim 1; Page 230-231; 2979pp; Japanese.
XX The invention relates to novel genes (AAI93926-AAI97963) expressed in
XX human neuroblastoma. The nucleic acids are applicable as a probe or
XX primer in diagnosing the prognosis of human neuroblastoma, malignancy and
XX susceptibility indicators or tumour markers for anti-cancer agents. The
XX gene information for diagnosing prognosis is related to factors similar
XX to that for N-myc and TrkA genes.
XX Sequence 889 BP; 240 A; 157 C; 142 G; 317 T; 33 other;

Query Match 0.5%; Score 19; DB 22; Length 889;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 627 TTTCATATGAAATATATG 645
DB 533 TTTCATATGAAATATATG 551

RESULT 35
AAU05849
ID AAX05849 standard; cDNA; 1296 BP.
XX
AC AAX05849;
XX 04-MAY-1999 (first entry)
XX Mouse pheromone receptor VR10 coding sequence.
XX Pheromone receptor; signal transduction; fertility; behaviour;
XX reproduction; rodent; insect; mouse; ss.
XX Mus sp.
XX WO9900422-A1.
XX 07-JAN-1999.
XX 30-JUN-1998; 98WO-US13680.
XX 30-JUN-1997; 97US-0051284.
XX (HARD) HARVARD COLLEGE.
```

XX PI Buck L, Dulac C, Herrada G, Matsunami H;  
 XX WPI: 1999-095684/08.  
 DR P-PSDB; AAW94910.  
 XX  
 PT New isolated pheromone receptor polypeptides - used to develop  
 PT products for controlling fertility and behaviour in vertebrates and  
 PT invertebrates  
 XX  
 PS Claim 23; Page 183; 308pp; English.  
 XX  
 CC The invention relates to polynucleotide sequences encoding mammalian  
 CC pheromone receptor polypeptides. The polypeptides are expressed in  
 CC murine and rat vomeronasal organ. The products can be used for modifying  
 CC pheromone activity, e.g. for decreasing pheromone receptor mediated  
 CC signal transduction. They can be used for controlling fertility and  
 CC behaviour in vertebrates and invertebrates. Compositions comprising the  
 CC polypeptides are particularly useful in e.g. controlling fertility in  
 CC livestock and controlling reproduction in rodents or insects by  
 CC interrupting the normal behaviours of rodents or insects that result in  
 CC reproduction. The present sequence represents the coding sequence of  
 CC mouse pheromone receptor VR10.  
 XX  
 SQ Sequence 1296 BP; 414 A; 229 C; 238 G; 415 T; 0 other;  
 Query Match 0.5%; Score 19; DB 20; Length 1296;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2806 TGATATGCCCATGAGTGAT 2824  
 Db 1152 TGATATGCCCATGAGTGAT 1170  
 RESULT 36  
 AAA61289  
 ID AAA61289 standard; DNA; 1405 BP.  
 AC AAA61289;  
 XX  
 DT 18-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein gene 30 clone HYBAR26.  
 XX  
 KW Human; secreted protein; fusion protein; gene therapy;  
 KW protein therapy; diagnosis; tissue; cancer; tumour; AIDS;  
 KW autoimmune disorder; allergy; cardiovascular; viral; bacterial;  
 KW fungal infection; immunosuppressive; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200029422-A1.  
 XX  
 PD 25-MAY-2000.  
 XX  
 XX 09-NOV-1999; 99WO-US26409.  
 PF  
 PR 12-NOV-1998; 98US-0108207.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Ni J, Ruben SM, Rosen CA, Ebner R, Florence KA, Young PE;  
 PI Birse CE, Carter KC, Komatsoulis G;  
 XX  
 XX WPI: 2000-387729/33.  
 DR  
 XX Novel human secreted proteins useful for diagnosing, preventing,  
 PT treating and ameliorating a medical condition e.g. cardiovascular  
 PT disease -  
 XX  
 XX Claim 1; Page 248; 295pp; English.  
 PS  
 XX

CC The present sequence represents a nucleic acid molecule which encodes a  
 CC secreted human protein. The gene number and the clone it was derived  
 CC from are given in the descriptor line.  
 CC The invention relates to 31 novel genes and their fragments (nucleic  
 CC acid sequences: AAA61260-A61293; amino acid sequences AAB12301-B12371)  
 CC which are useful for preventing, treating or ameliorating medical  
 CC conditions e.g. by protein or gene therapy. Also pathological conditions  
 CC can be diagnosed by determining the amount of the new polypeptides in a  
 CC sample or by determining the presence of mutations in the new  
 CC polynucleotides. Specific uses are described for each of the 31  
 CC polynucleotides, based on which tissues they are most highly expressed in  
 CC and include products for the diagnosis or treatment of cancer, tumours,  
 CC AIDS, autoimmune disorders, allergy, cardiovascular disorders, viral,  
 CC bacterial and fungal infection. The genes are used to generate fusion  
 CC proteins by linking to the gene a human immunoglobulin portion (AAA61251)  
 CC for increasing stability of the fused protein as compared to the  
 CC secreted protein only.  
 XX  
 SQ Sequence 1405 BP; 410 A; 264 C; 312 G; 412 T; 7 other;  
 Query Match 0.5%; Score 19; DB 21; Length 1405;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2300 ATGGGATGTTGGATGAAT 2318  
 Db 822 ATGGGATGTTGGATGAAT 840

RESULT 37  
 AXA05850  
 ID AAX05850 standard; cDNA; 1521 BP.  
 XX  
 AC AAX05850;  
 XX  
 DT 04-MAY-1999 (first entry)  
 XX  
 DE Mouse pheromone receptor VR11 coding sequence.  
 XX  
 KW Pheromone receptor; signal transduction; fertility; behaviour;  
 KW reproduction; rodent; insect; mouse; ss.  
 XX  
 OS Mus sp.  
 XX  
 PN WO9900422-A1.  
 XX  
 PD 07-JAN-1999.  
 PF  
 XX 30-JUN-1998; 98WO-US13680.  
 XX  
 PR 30-JUN-1997; 97US-0051284.  
 XX  
 PA (HARD ) HARVARD COLLEGE.  
 XX  
 PI Buck L, Dulac C, Herrada G, Matsunami H;  
 XX  
 DR WPI: 1999-095684/08.  
 DR P-PSDB; AAW94911.  
 XX  
 PT New isolated pheromone receptor polypeptides - used to develop  
 PT products for controlling fertility and behaviour in vertebrates and  
 PT invertebrates  
 XX  
 PS Claim 23; Page 184; 308pp; English.  
 XX  
 CC The invention relates to polynucleotide sequences encoding mammalian  
 CC pheromone receptor polypeptides. The polypeptides are expressed in  
 CC murine and rat vomeronasal organ. The products can be used for modifying  
 CC pheromone activity, e.g. for decreasing pheromone receptor mediated  
 CC signal transduction. They can be used for controlling fertility and  
 CC behaviour in vertebrates and invertebrates. Compositions comprising the  
 CC polypeptides are particularly useful in e.g. controlling fertility in  
 CC livestock and controlling reproduction in rodents or insects by

CC interrupting the normal behaviours of rodents or insects that result in  
CC reproduction. The present sequence represents the coding sequence of  
CC mouse pheromone receptor VR10.  
XX  
SQ Sequence 1521 BP; 489 A; 264 C; 288 G; 480 T; 0 other;  
Query Match 0.5%; Score 19; DB 20; Length 1521;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2806 TGATATGCCCATGAGTGAT 2824  
|||||  
Db 1152 TGATATGCCCATGAGTGAT 1170  
RESULT 38  
ABL08786  
ID ABL08786 standard; cDNA; 2606 BP.  
XX  
AC ABL08786;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 20840.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WC200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR P-PSDB; ABB64683.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1; SEQ ID NO 20840; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2606 BP; 760 A; 496 C; 506 G; 844 T; 0 other;  
Query Match 0.5%; Score 19; DB 23; Length 2606;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2262 CTAGATCAAGATATTAT 2280  
|||||  
Db 338 CTAGATCAAGATATTAT 356

us-09-782-874-1.oli.rng

RESULT 39  
AA05810  
ID AA05810 standard; cDNA; 2732 BP.  
XX  
AC AA05810;  
XX  
DT 04-MAY-1999 (first entry)  
XX  
DE Mouse pheromone receptor VR10 encoding cDNA.  
XX  
KW Pheromone receptor; signal transduction; fertility; behaviour;  
KW reproduction; rodent; insect; mouse; ss.  
XX  
OS Mus sp.  
XX  
FH Key Location/Qualifiers  
FT CDS 80..1378  
FT /\*tag= a  
XX  
PN WO9900422-A1.  
XX  
PD 07-JAN-1999.  
XX  
PF 30-JUN-1998; 98WO-US13680.  
XX  
PR 30-JUN-1997; 97US-0051284.  
XX  
PA (HARD ) HARVARD COLLEGE.  
XX  
PI Buck L, Dulac C, Herrada G, Matsunami H;  
XX  
DR WPI; 1999-095684/08.  
DR P-PSDB; AAW94910.  
XX  
PT New isolated pheromone receptor polypeptides - used to develop  
PT products for controlling fertility and behaviour in vertebrates and  
PT invertebrates  
XX  
PS Claim 18; Page 99-101; 308pp; English.  
XX  
CC The invention relates to polynucleotide sequences encoding mammalian  
CC pheromone receptor polypeptides. The polypeptides are expressed in  
CC murine and rat vomeronasal organ. The products can be used for modifying  
CC pheromone activity, e.g. for decreasing pheromone receptor mediated  
CC signal transduction. They can be used for controlling fertility and  
CC behaviour in vertebrates and invertebrates. Compositions comprising the  
CC polypeptides are particularly useful in e.g. controlling fertility in  
CC livestock and controlling reproduction in rodents or insects by  
CC interrupting the normal behaviours of rodents or insects that result in  
CC reproduction. The present sequence represents a cDNA encoding a mouse  
CC pheromone receptor VR10. The cDNA is deposited under the Genbank  
CC accession number AF011420.  
XX  
SQ Sequence 2732 BP; 818 A; 532 C; 509 G; 873 T; 0 other;  
Query Match 0.5%; Score 19; DB 20; Length 2732;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2806 TGATATGCCCATGAGTGAT 2824  
|||||  
Db 1231 TGATATGCCCATGAGTGAT 1249  
RESULT 40  
AA05811  
ID AA05811 standard; cDNA; 2962 BP.  
XX  
AC AA05811;  
XX  
DT 04-MAY-1999 (first entry)

```

XX Mouse pheromone receptor VR11 encoding cDNA.
DE
XX
XX Pheromone receptor; signal transduction; fertility; behaviour;
KW reproduction; rodent; insect; mouse; ss.
XX
XX Mus sp.
OS
XX
XX Key Location/Qualifiers
FH 81..1604
FT CDS /*tag= a
FT
XX WO9900422-A1.
XX
XX 07-JAN-1999.
XX
XX 30-JUN-1998; 98WO-US13680.
XX
XX 30-JUN-1997; 97US-0051284.
XX (HARD ) HARVARD COLLEGE.
XX
XX Buck L, Dulac C, Herrada G, Matsunami H;
PI WPI; 1999-095684/08.
DR P-PSDB; AAW94911.
XX
XX New isolated pheromone receptor polypeptides - used to develop
PT products for controlling fertility and behaviour in vertebrates and
PT invertebrates
XX
XX Claim 18; Page 102-105; 308pp; English.
XX
XX The invention relates to polynucleotide sequences encoding mammalian
CC pheromone receptor polypeptides. The polypeptides are expressed in
CC murine and rat vomeronasal organ. The products can be used for modifying
CC pheromone activity, e.g. for decreasing pheromone receptor mediated
CC signal transduction. They can be used for controlling fertility and
CC behaviour in vertebrates and invertebrates. Compositions comprising the
CC polypeptides are particularly useful in e.g. controlling fertility in
CC livestock and controlling reproduction in rodents or insects by
CC interrupting the normal behaviours of rodents or insects that result in
CC reproduction. The present sequence represents a cDNA encoding a mouse
CC pheromone receptor VR11. The cDNA is deposited under the Genbank
CC accession number AF011421.
XX
XX Sequence 2962 BP; 898 A; 567 C; 559 G; 938 T; 0 other;
SQ
Query Match 0.5%; Score 19; DB 20; Length 2962;
Best Local Similarity 100.0%; Pred. NO. 1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2806 TGATATGCCCATGAGTGAT 2824
DB 1232 TGATATGCCCATGAGTGAT 1250
|||||
RESULT 41
AAV36034/C
ID AAV36034 standard; DNA; 4446 BP.
XX
XX AAV36034;
AC
XX 02-SEP-1998 (first entry)
DT
XX DNA encoding a DNA polymerase of Pyrococcus sp. GE23.
DE
XX Heat stable DNA polymerase; archaeobacteria; secondary structure; ds.
KW
XX Pyrococcus sp.
OS
XX Key Location/Qualifiers
FH 1547..3862
FT CDS

```

---

```

FT /*tag= a
XX /product= DNA polymerase
XX
XX WO9801567-A2.
XX
XX 15-JAN-1998.
XX
XX 10-JUL-1997; 97WO-FR01259.
XX
XX 10-JUL-1996; 96FR-0008631.
XX
XX (APPL-) APPLIGENE-ONCOR SA.
XX
XX Cambon MA, Querellou J;
PI WPI; 1998-101062/09.
DR P-PSDB; AAW60719.
XX
XX Heat stable DNA polymerase from Pyrococcus species - and related
PT DNA, vectors and transformed cells, useful in nucleic acid
PT amplification reactions performed at temperatures high enough to
PT denature secondary structures
XX
XX Claim 9; Pages 37-41; 80pp; French.
XX
XX The present sequence encodes a heat stable DNA polymerase from
CC archaeobacteria of the genus Pyrococcus sp. GE23. The encoded polymerase
CC has an extremely high heat stability and can catalyse reactions at
CC 70-90 degrees Celsius. High molecular weight DNA was isolated from the
CC bacteria, and screened with probes based on the polymerase genes of
CC P. furiosus and Thermococcus litoralis to isolate the present sequence.
CC Host cells transformed with DNA encoding the present polymerase are
CC used to produce recombinant DNA which is useful for nucleic acid
CC amplification, including direct, reversed polymerase chain reaction (PCR)
CC on double-stranded matrices and PCR on matrices having secondary
CC structures that block the process at conventional elongation
CC temperatures.
XX
XX Sequence 4446 BP; 1358 A; 884 C; 1115 G; 1089 T; 0 other;
SQ
Query Match 0.5%; Score 19; DB 19; Length 4446;
Best Local Similarity 100.0%; Pred. NO. 1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 221 TTCCTTATCTCTCTCTG 239
DB 1852 TTCCTTATCTCTCTCTG 1834
|||||
RESULT 42
AAL36325
ID AAL36325 standard; DNA; 4656 BP.
XX
XX AAL36325;
AC
XX 08-JAN-2002 (first entry)
DT
XX Human musculoskeletal system related polynucleotide SEQ ID NO 2690.
DE
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; anti-ulcer;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ds.
XX
XX Homo sapiens.
OS
XX WO200155367-A1.
PN
XX 02-AUG-2001.
PD
XX 17-JAN-2001; 2001WO-US01338.
PF

```

XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0215647.  
PR 07-JUL-2000; 2000US-0215880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225477.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231142.  
PR 08-SEP-2000; 2000US-0231143.  
PR 08-SEP-2000; 2000US-0231144.  
PR 08-SEP-2000; 2000US-0231143.  
PR 08-SEP-2000; 2000US-0231144.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234397.  
PR 25-SEP-2000; 2000US-0234398.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235836.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0248526.  
PR 08-NOV-2000; 2000US-0248527.  
PR 08-NOV-2000; 2000US-0248528.  
PR 08-NOV-2000; 2000US-0248532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
(HUMA-) HUMAN GENOME SCI INC.  
PA Rosen CA, Barash SC, Ruben SM;  
PI WPI; 2001-451937/48.  
XX Isolated polypeptide for treating, preventing and/ or prognosing  
XX disorders related to the musculoskeletal system including  
PT

PT musculoskeletal cancers and also for testing and detection e.g.  
 PT diagnosis -  
 XX  
 PS Example 2: SEQ ID NO 2690; 781pp + Sequence Listing; English.  
 XX  
 CC The invention relates to novel genes (AAL34669-AAL37666) and proteins  
 CC (ABB03087-ABB04109) associated with the musculoskeletal system useful  
 CC for preventing, treating or ameliorating medical conditions e.g. by  
 CC protein or gene therapy. The genes are isolated from a range of human  
 CC tissues disclosed in the specification. The nucleic acids, proteins,  
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment  
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and  
 CC other cancers of the adrenal gland, bone, bone marrow, breast,  
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune  
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;  
 CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound  
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;  
 CC and (f) infectious diseases such as viral, bacterial, fungal and  
 CC parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 4656 BP; 1461 A; 810 C; 804 G; 1581 T; 0 other;

Query Match 0.5%; Score 19; DB 22; Length 4656;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3567 TATATAATGTAATAGGGTG 3585  
 |||||  
 DB 2822 TATATAATGTAATAGGGTG 2840

RESULT 43  
 AAD11120/c  
 ID AAD11120 standard; DNA; 4877 BP.  
 AC AAD11120;  
 XX  
 XX 24-SEP-2001 (first entry)  
 DT  
 DE Human small cell lung cancer associated gene, KIAA0963.  
 XX  
 KW Human; small cell lung cancer; therapy; hCAAP; nucleic acid; NA;  
 KW melanoma; cancer; colon; breast; head; neck; transitional cancer;  
 KW leiomyosarcoma; synovial sarcoma; cytostatic; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153349-A2.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 19-JAN-2001; 2001WO-US02015.  
 XX  
 PR 21-JAN-2000; 2000US-0489101.  
 XX  
 XX (LUDW-) LUDWIG INST CANCER RES.  
 PA (SLOK) SLOAN KETTERING INST CANCER RES.  
 PA (CORR) CORNELL RES FOUND INC.  
 XX  
 PI Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;  
 XX  
 DR WPI; 2001-457597/49.  
 XX  
 PT Isolated polypeptide, used to treat or prognose a disorder  
 PT characterized by expression of a hCAAP e.g. cancer, is encoded by an  
 PT isolated nucleic acid comprising an NA Group 3 or 4 molecule -  
 XX  
 PS Claim 57; Page 105-107; 152pp; English.

XX The invention relates to nucleic acids and encoded polypeptides which  
 CC are cancer associated antigens expressed in patients afflicted with  
 CC small cell lung cancer. The molecules provided by the invention can be  
 CC used in the diagnosis, monitoring, research or treatment of conditions  
 CC characterised by the expression of one or more cancer associated  
 CC antigens. The polypeptide is used to treat a disorder characterised by  
 CC expression of a hCAAP, and determine regression, progression or onset  
 CC of a condition characterised by expression of an abnormal amount of a  
 CC protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders  
 CC are small and non-small cell lung cancer, melanoma, colon, breast, head  
 CC and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.  
 CC The present sequence is a small cell lung cancer associated gene  
 CC designated as NY-SCLC-11, encoding human KIAA0963 protein.  
 XX  
 SQ Sequence 4877 BP; 867 A; 1682 C; 1570 G; 758 T; 0 other;

Query Match 0.5%; Score 19; DB 22; Length 4877;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2775 ATGCCCATGTCGATTGTC 2793  
 |||||  
 DB 3238 ATGCCCATGTCGATTGTC 3220

RESULT 44  
 AAC75998/c  
 ID AAC75998 standard; cDNA; 4924 BP.  
 XX AAC75998;  
 AC AAC75998;  
 XX  
 XX 08-FEB-2001 (first entry)  
 DT  
 DE Human ORFX ORF1553 polynucleotide sequence SEQ ID NO:3105.  
 XX  
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnery; antipsoriatic; antiparkinsonian; nocrotic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200058473-A2.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 31-MAR-2000; 2000WO-US08621.  
 XX  
 PR 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 PA  
 XX Shimkets RA, Leach M;  
 PI  
 DR WPI; 2000-602362/57.  
 DR P-PSDB; AAB41789.  
 XX  
 XX Novel nucleic acids and peptides derived from open reading frame x,  
 PT useful for treating e.g. cancers, proliferative disorders,



PT neurodegenerative disorders and cardiovascular disease -  
 XX Claim 5; Page 2323-2326; 5507pp; English.  
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;  
 CC antiparasitic; antiparkinsonian; nootropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX Sequence 4924 BP; 892 A; 1694 C; 1572 G; 765 T; 1 other;  
 XX

Query Match 0.5%; Score 19; DB 21; Length 4924;  
 Best Local Similarity 100.0%; Pred. No. 1e-02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2775 ATGCCCATGTCGTATTTC 2793  
 Db 3239 ATGCCCATGTCGTATTTC 3221  
 |||||

RESULT 45  
 ABL30262/c  
 ID ABL30262 standard; DNA; 5415 BP.  
 XX  
 AC ABL30262;  
 XX

DT 26-MAR-2002 (first entry)  
 XX

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 42259.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX

OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX

PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 XX (PEKE ) PE CORP NY.  
 PA

PI Venter JC, Adams M, Li PWD, Myers EW;  
 PI WPI; 2001-656860/75.  
 DR

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PT  
 PS Claim 1; SEQ ID NO 42259; 21pp + Sequence Listing; English.  
 XX

CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (AB57737-AB572072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 5415 BP; 1512 A; 1457 C; 1267 G; 1179 T; 0 other;  
 XX

Query Match 0.5%; Score 19; DB 23; Length 5415;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3562 ATATATATATAATGTAATA 3580  
 Db 5286 ATATATATATAATGTAATA 5268  
 |||||

RESULT 46  
 AAH62719  
 ID AAH62719 standard; DNA; 18234 BP.  
 XX  
 AC AAH62719;  
 XX

DT 11-SEP-2001 (first entry)  
 XX

DE Shrimp white spot Bacilliform virus (WSBV) gene 30.  
 XX  
 KW Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;  
 KW antiviral agent; gene expression; antisense construct;  
 KW transgenic viral resistant shrimp; ds.  
 XX

OS White spot syndrome virus.  
 XX  
 PN WO200138351-A2.  
 XX  
 PD 31-MAY-2001.  
 XX

PF 08-NOV-2000; 2000WO-US28888.  
 XX  
 XX 24-NOV-1999; 99CN-0124717.  
 XX

XX (PENY-) PE CORP NY.  
 PA  
 PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.  
 PA (SINO-) SINOGENOMAX CO LTD.  
 XX

PI Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;  
 PI WPI; 2001-355877/37.  
 DR  
 DR P-PSDB; AAG84939.  
 XX

XX Primary nucleotide sequence of the shrimp white spot Bacilliform virus  
 PT (WSBV), useful for producing viral polypeptides that can be used to  
 PT screen for agents that are useful for treating WSBV infection -  
 XX  
 PS Claim 4; Figure 3; 626pp; English.  
 XX

CC The invention provides the primary nucleotide sequence of the WSBV genome  
 CC (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and  
 CC encoded proteins (AAG84910-AAG85031) and oligonucleotide sequences  
 CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid  
 CC molecules and proteins of the invention are useful for diagnosis and  
 CC monitoring viral infection, in screens for antiviral agents and for  
 CC monitoring viral gene expression or activity during a treatment regimen.  
 CC The nucleic acid molecules are also useful as antisense constructs to  
 CC control viral gene expression in infected cells and tissues and to create  
 CC transgenic viral resistant shrimp.  
 XX

```

SQ Sequence 18234 BP; 5373 A; 4270 C; 3810 G; 4781 T; 0 other;
Query Match 0.5%; Score 19; DB 22; Length 18234;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 554 ATTCAATTGGCATTGGAT 572
Db 4662 ATTCAATTGGCATTGGAT 4680

RESULT 47
AAH62689
ID AAH62689 standard; DNA; 305107 BP.
XX
AC AAH62689;
XX
DT 11-SEP-2001 (first entry)
XX
DE Shrimp white spot Bacilliform virus (WSBV) genomic sequence.
XX
KW Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
KW antiviral agent; gene expression; antisense construct;
KW transgenic viral resistant shrimp; ds.
XX
OS White spot syndrome virus.
XX
PN WO200138351-A2.
XX
PD 31-MAY-2001.
XX
PF 08-NOV-2000; 2000WO-US28888.
XX
PR 24-NOV-1999; 99CN-0124717.
XX
PA (PENY-) PE CORP NY.
PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
PA (SINO-) SINOGENOMAX CO LTD.
XX
PI Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;
XX WPI; 2001-355877/37.
XX
PT Primary nucleotide sequence of the shrimp white spot Bacilliform virus
PT (WSBV), useful for producing viral polypeptides that can be used to
PT screen for agents that are useful for treating WSBV infection -
XX
PS Disclosure; Figure 2; 626pp; English.
XX
CC The invention provides the primary nucleotide sequence of the WSBV genome
CC (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and
CC encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences
CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
CC molecules and proteins of the invention are useful for diagnosis and
CC monitoring viral infection, in screens for antiviral agents and for
CC monitoring viral gene expression or activity during a treatment regimen.
CC The nucleic acid molecules are also useful as antisense constructs to
CC control viral gene expression in infected cells and tissues and to create
CC transgenic viral resistant shrimp.
XX
SQ Sequence 305107 BP; 92042 A; 62482 C; 62635 G; 87948 T; 0 other;
Query Match 0.58; Score 19; DB 22; Length 305107;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 554 ATTCAATTGGCATTGGAT 572
Db 214277 ATTCAATTGGCATTGGAT 214295

RESULT 48
AAD04376/c

```

---

```

ID AAD04376 standard; DNA; 35 BP.
XX
AC AAD04376;
XX
DT 04-JUL-2001 (first entry)
XX
DE Reverse PCR primer, P127BGL to amplify tomato RdRP cDNA.
XX
KW Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
KW transgenic plant; tissue culture; plant breeding; therapy; C-protein;
KW PCR primer; ss.
XX
OS Lycopersicon esculentum.
XX
PN US6218142-B1.
XX
PD 17-APR-2001.
XX
PF 05-MAR-1997; 97US-0811583.
XX
PR 05-MAR-1997; 97US-0811583.
XX
PA (WASS/) WASENEGGER M.
PA (RIED/) RIEDEL L.
XX
PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX WPI; 2001-289830/30.
XX
PT New nucleic acid molecules encoding polypeptides with RNA-directed RNA
PT polymerase enzymatic activity, useful in modulating gene expression in
PT plants, humans and animals, as well as in plant cell/tissue cultures or
PT plant breeding -
XX
PS Example 3; Column 21; 31pp; English.
XX
CC The present sequence is reverse PCR primer, P127BGL, used to amplify a
CC cDNA encoding tomato C-protein having RNA-directed RNA polymerase (RdRP)
CC activity. This primer contains BglII site to enhance the cloning
CC efficiency. This protein is capable of RNA-directed RNA synthesis, thus
CC using RNA as a template for synthesizing complementary RNA molecules.
CC RdRP nucleic acid is useful for modulating gene expression in plants,
CC humans and animals. This may lead to various physiological, developmental
CC and/or morphological changes. Transgenic plants containing RdRP nucleic
CC acid is especially useful in plant cell or tissue cultures and in plant
CC breeding. RdRP is useful in gene therapy, particularly for treating a
CC disease that is caused by the undesirable expression or overexpression of
CC a gene.
XX
SQ Sequence 35 BP; 13 A; 7 C; 8 G; 7 T; 0 other;
Query Match 0.5%; Score 18; DB 22; Length 35;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3615 GACTACCTTTGCTTTA 3632
Db 35 GACTACCTTTGCTTTA 18

RESULT 49
AAS17843/c
ID AAS17843 standard; DNA; 35 BP.
XX
AC AAS17843;
XX
DT 08-MAY-2002 (first entry)
XX
DE Tomato RNA-directed RNA polymerase (RdRP) P127Bgl PCR primer.
XX
KW Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
KW cytosstatic; virucide; RNA synthesis inhibitor; antibody; immunogen;
KW transgenic plant; transgenic animal; cancer; viral infection; RACE;

```

KW immunoprecipitation; immunolocalisation; PCR primer; p127Bgl;  
KW gene therapy; ss.

XX Lycopersicon esculentum.

XX US2001023067-A1.

XX 20-SEP-2001.

XX 08-FEB-2001; 2001US-0782874.

XX 05-MAR-1997; 97US-0811583.

XX (WASS/) WASSENEGGER M.

XX (RIED/) RIEDEL L.

XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;

XX WPI; 2001-595798/67.

XX New nucleic acid molecule encoding a polypeptide having the enzymatic  
XX activity of RNA-directed RNA polymerase, for modulating gene expression  
XX and treating cancer and virus infection in human and animals

XX Example 3; Page 12; 34pp; English.

XX This sequence represents the tomato RNA-directed RNA polymerase (RdRP)  
XX specific PCR primer p127Bgl used to clone the full length RdRP cDNA  
XX molecule of the invention. The invention comprises the nucleic acid and  
XX protein sequences of RdRP, the protein of the invention can catalyze in  
XX vitro transcription of short single stranded RNAs into DNA molecules,  
XX this transcription can be either primed by RNA or DNA oligonucleotides,  
XX or be unprimed. The protein may have cytotatic or virucide activities.  
XX The sequences of the invention may be used in gene therapy or as an RNA  
XX directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template  
XX nucleic acid molecule derived from a nucleic acid molecule which causes  
XX a disease are useful for treating a disease caused by the undesired  
XX expression or overexpression of a nucleic acid molecule in a human, rat  
XX or mouse, by administering the molecules. This system can be used in the  
XX preparation of a pharmaceutical composition and for inhibiting to  
XX expression of any desired gene by transferring the RdRP system to  
XX organisms that either lack a comparable mechanism or do not sufficiently  
XX express their own RdRP. An antibody or an antagonist or inhibitor to the  
XX protein are useful for inhibiting RNA directed RNA synthesis and for  
XX ensuring stable heterologous, gene expression in transgenic organisms.  
XX The sequence is useful for probes and/or for the control of gene  
XX expression, as primers for amplification of nucleic acid molecules and  
XX as tools for the detection of expression of the cDNA molecules.  
XX Additionally, nucleotide and protein sequences are useful for  
XX suppression of undesired gene expression in humans and animals. The RdRP  
XX is useful as a therapeutic agent for the control of cancer and virus  
XX infection in humans and animals and the antibody is useful for  
XX immunoprecipitation or immunolocalisation of the protein, identification  
XX of polypeptides interacting with it and screening expression libraries.

XX Sequence 35 BP; 13 A; 7 C; 8 G; 7 T; 0 other;

Query Match 0.5%; Score 18; DB 23; Length 35;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3615 GACTACCTTTGCTCTTA 3632  
DB 35 GACTACCTTTGCTCTTA 18  
|||||

RESULT 50  
RAE29312  
ID AAF29312 standard; DNA; 48 BP.  
XX AC  
XX AAF29312;  
XX 18-APR-2001 (first entry)  
DT

XX

DE Primer base sequence used to illustrate primer selection method.

XX

KW Primer; optimum sequence; differential display; ss.

XX

OS Synthetic.

XX

PN JP2000308487-A.

XX

PD 07-NOV-2000.

XX

XX 30-MAR-1999; 95JP-0088410.

PF

XX 30-MAR-1999; 95JP-0088410.

PR

XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

PA

XX WPI; 2001-046077/06.

DR

XX Selection of primer base for optimizing primer selection comprises  
PT obtaining an optimum sequence for differential display from an  
PT expression gene data base

XX

PS Disclosure; Fig 9; 13pp; Japanese.

XX

CC This invention relates to a method for selecting the sequence of a  
CC primer. The method comprises obtaining an optimum sequence for  
CC differential display from an expression gene data base, and using the  
CC base sequences most frequently expressed as the primer candidates in the  
CC order of frequency. The optimum primer group characterised by the use of  
CC genetic algorithm from the primer candidates is selected. The method is  
CC used for selecting a primer sequence quickly. The present sequence  
CC represents a primer used in an illustration of the method of the  
CC invention.

XX

SQ Sequence 48 BP; 18 A; 2 C; 5 G; 23 T; 0 other;

Query Match

Best Local Similarity 0.5%; Score 18; DB 22; Length 48;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 571 ATTGAAGAAATTTT 588

|||||

DB 5 ATTGAAGAAATTTT 22

Search completed: November 5, 2002, 23:52:42  
Job time : 1690 secs



FEATURE:  
NAME/KEY: CDS  
LOCATION: 194...3535  
US-08-811-583-1

Query Match 100.0%; Score 3731; DB 4; Length 3731;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAATATTCCTTACTTACTTACACAGGATGACTCATCTACCTCCCTCAAGTCTTTGTGT 60  
DB 1 GAAATATTCCTTACTTACTTACACAGGATGACTCATCTACCTCCCTCAAGTCTTTGTGT 60

QY 61 GTTGTGATATAAATTTGGTTGGCTTCAGTTTCAGTCACTACTGCTGGGTAGTTTTAT 120  
DB 61 GTTGTGATATAAATTTGGTTGGCTTCAGTTTCAGTCACTACTGCTGGGTAGTTTTAT 120

QY 121 TTTGCTAACTTCAGGGGATTCAGTTGGTTAGCATTTGAAAGTCGAAGTGCACAT 180  
DB 121 TTTGCTAACTTCAGGGGATTCAGTTGGTTAGCATTTGAAAGTCGAAGTGCACAT 180

QY 181 GGAATTTGGCTACATGGGAAAGACAATTCAGGTTTTCGGATTCCTTCTCTCTGC 240  
DB 181 GGAATTTGGCTACATGGGAAAGACAATTCAGGTTTTCGGATTCCTTCTCTCTGC 240

QY 241 GGAAGTGGTTAAGTCATCTTACAGAAATATACAGATATGAACGTATGTGCATTGA 300  
DB 241 GGAAGTGGTTAAGTCATCTTACAGAAATATACAGATATGAACGTATGTGCATTGA 300

QY 301 GGTAAACAGTCCAAAGGAGATCTAGACATTTGCCAAGTTCAATTTGCCGACAAAT 360  
DB 301 GGTAAACAGTCCAAAGGAGATCTAGACATTTGCCAAGTTCAATTTGCCGACAAAT 360

QY 361 AAGTGTGACAAATCATCTACTTGGCTAATACAGGCTGTATTTGGCTCTCTCTTTT 420  
DB 361 AAGTGTGACAAATCATCTACTTGGCTAATACAGGCTGTATTTGGCTCTCTCTTTT 420

QY 421 GAAGCTTGGGAAATGAAACTGATATGTCCAACTGCGGCAATATGTGGATCAGATGA 480  
DB 421 GAAGCTTGGGAAATGAAACTGATATGTCCAACTGCGGCAATATGTGGATCAGATGA 480

QY 481 TGGCATTAATTTGAAATTCGGATTCAGATATCAGATGACAACTTTGCAGTGTGGGAG 540  
DB 481 TGGCATTAATTTGAAATTCGGATTCAGATATCAGATGACAACTTTGCAGTGTGGGAG 540

QY 541 TACAGAGTTTCAATTCAAATTTGGCATTTGATTTGAAGAAATTTTTTTCTTTTATCTAG 600  
DB 541 TACAGAGTTTCAATTCAAATTTGGCATTTGATTTGAAGAAATTTTTTTCTTTTATCTAG 600

QY 601 TGGTTCAGTCACTATAAATTCAGCTTTTCATATGAAATATATATGCGAGTTTGTCTCCA 660  
DB 601 TGGTTCAGTCACTATAAATTCAGCTTTTCATATGAAATATATATGCGAGTTTGTCTCCA 660

QY 661 TCGTCCATATGGTCAAAATGCTCAGTTTCTCCTCATACAGTTATTTGGTCTCTCTCGGAT 720  
DB 661 TCGTCCATATGGTCAAAATGCTCAGTTTCTCCTCATACAGTTATTTGGTCTCTCTCGGAT 720

QY 721 CTATAAGAGACTTGAAGAACTCCCTGTTATAGCTTCTTTAAGGAACTCCTGATCAGTG 780  
DB 721 CTATAAGAGACTTGAAGAACTCCCTGTTATAGCTTCTTTAAGGAACTCCTGATCAGTG 780

QY 781 GGTGAGGACAAAGATTTCCCTCCATCTGATAGGCTATCTTCAGCTTTATGTTTGA 840  
DB 781 GGTGAGGACAAAGATTTCCCTCCATCTGATAGGCTATCTTCAGCTTTATGTTTGA 840

QY 841 GTTCCGTAGGGGTGCTGCTTCCAAATTTTCAGGAAAGTTTTTCCACTATGCAAGC 900  
DB 841 GTTCCGTAGGGGTGCTGCTTCCAAATTTTCAGGAAAGTTTTTCCACTATGCAAGC 900

QY 901 TGAAACAATATTAATTTACAGCTGTTTCACTTTTGGTCTCTCAAAAATTCGGCTCT 960  
DB 901 TGAAACAATATTAATTTACAGCTGTTTCACTTTTGGTCTCTCAAAAATTCGGCTCT 960

QY 961 GGTTCCTCCCAATGTCCAGCCTCCGGAAGAAATTTCAATTCCTTACAGATTTTGTTCAAAAT 1020  
DB 961 GGTTCCTCCCAATGTCCAGCCTCCGGAAGAAATTTCAATTCCTTACAGATTTTGTTCAAAAT 1020

QY 1021 TAGTTCTTTTGTACAGCATGGATGATACCTGGCCAGCATTAATGATGCTACTTTTTCCG 1080  
DB 1021 TAGTTCTTTTGTACAGCATGGATGATACCTGGCCAGCATTAATGATGCTACTTTTTCCG 1080

QY 1081 ATTAGTTGATCCTCGAAGGAGAAATGTGGCATGTGATGATGAGCATGCTTTAGAGAACTGTA 1140  
DB 1081 ATTAGTTGATCCTCGAAGGAGAAATGTGGCATGTGATGATGAGCATGCTTTAGAGAACTGTA 1140

QY 1141 CTATATAAGGAGTCTGTTATGATCCCGTGAGTGGCTCACTGAGCAGTATGATGGTGA 1200  
DB 1141 CTATATAAGGAGTCTGTTATGATCCCGTGAGTGGCTCACTGAGCAGTATGATGGTGA 1200

QY 1201 TCTCAAGGAGTACAAACCTCCAAAATCTCCGTCCATCACTTTAGATGATGGTTGGTGA 1260  
DB 1201 TCTCAAGGAGTACAAACCTCCAAAATCTCCGTCCATCACTTTAGATGATGGTTGGTGA 1260

QY 1261 TGTAAAGAGGTCCTAGTAACACCATGCAAAAGTTTATTTTGTGTCAGAGGTTAATGT 1320  
DB 1261 TGTAAAGAGGTCCTAGTAACACCATGCAAAAGTTTATTTTGTGTCAGAGGTTAATGT 1320

QY 1321 TTCCCAATCGGTTCTCCGCAATTTATCTGAAGACATAGATAAATTTCTTCTGTTCTTT 1380  
DB 1321 TTCCCAATCGGTTCTCCGCAATTTATCTGAAGACATAGATAAATTTCTTCTGTTCTTT 1380

QY 1381 TGTGATGAGGAGTGGGAGAACTGTATCTACAGACTTATACCAAGCAAGTACTGG 1440  
DB 1381 TGTGATGAGGAGTGGGAGAACTGTATCTACAGACTTATACCAAGCAAGTACTGG 1440

QY 1441 AAGTGGTGTACAGACAAACATCTATGAGAGATCTTATCAACTCTGCGAAAGGCTTTGT 1500  
DB 1441 AAGTGGTGTACAGACAAACATCTATGAGAGATCTTATCAACTCTGCGAAAGGCTTTGT 1500

QY 1501 AATTTGGTGAATAAATTTGAAATTTCTTGCATTTTCATCGAGCCAGTTGCGGGATATTC 1560  
DB 1501 AATTTGGTGAATAAATTTGAAATTTCTTGCATTTTCATCGAGCCAGTTGCGGGATATTC 1560

QY 1561 AGTGTGGATTTTGCATCAAGACCTGGCTTTACTGCAATGATATAAGAGCTTTGATGGG 1620  
DB 1561 AGTGTGGATTTTGCATCAAGACCTGGCTTTACTGCAATGATATAAGAGCTTTGATGGG 1620

QY 1621 TGATTTTTCGAGATCAAGATGTCGAAAATATGCTGCCAGACTTGGTCAATCTTTTGG 1680  
DB 1621 TGATTTTTCGAGATCAAGATGTCGAAAATATGCTGCCAGACTTGGTCAATCTTTTGG 1680

QY 1681 TTCTCCAGAGAGCTTTGAGTGTCTTAGGCATGAGATTGAAGTTATCCCGATGATAA 1740  
DB 1681 TTCTCCAGAGAGCTTTGAGTGTCTTAGGCATGAGATTGAAGTTATCCCGATGATAA 1740

QY 1741 GGTTCATGGAACCCAGCTATGCTTTTCTGATGGAATTTGTAATAATCTGTTGACTTTGC 1800  
DB 1741 GGTTCATGGAACCCAGCTATGCTTTTCTGATGGAATTTGTAATAATCTGTTGACTTTGC 1800

QY 1801 TCATAGAGTTGCTCAAAATGTCGCTTCAATATACCCCATCTGTTTCCAGATTCGTTA 1860  
DB 1801 TCATAGAGTTGCTCAAAATGTCGCTTCAATATACCCCATCTGTTTCCAGATTCGTTA 1860

QY 1861 TGGTGGATATAAAGTGTGTTGGTGTGATCCGGATTCATCAATGAAGTTGCTCTTTGAG 1920  
DB 1861 TGGTGGATATAAAGTGTGTTGGTGTGATCCGGATTCATCAATGAAGTTGCTCTTTGAG 1920

QY 1921 AAAGAGCATGCGAAATATGATGATCAGACAAATAGTAGATGCTTGGATGAGGAA 1980  
DB 1921 AAAGAGCATGCGAAATATGATGATCAGACAAATAGTAGATGCTTGGATGAGGAA 1980

QY 1981 ATATCAGCCTTCTTATCTTAATCGTCACTGATTAACGTTCTTGTCTACACTTTGGAGTGA 2040  
DB 1981 ATATCAGCCTTCTTATCTTAATCGTCACTGATTAACGTTCTTGTCTACACTTTGGAGTGA 2040

QY 2041 AGATGAAGTTCTCGAACAGCAAAAGGAGCTGTAGATCAGCTTGATCTATCTTGA 2100



REFERENCE/DOCKET NUMBER: 7555-00004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (313) 641-1600  
TELEFAX: (313) 641-0270  
TELEX: 287637 Harness UR  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 906 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Solanum tuberosum  
INDIVIDUAL ISOLATE: p-PCM-1  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 81..530  
US-08-100-874-1

Query Match 0.8%; Score 30; DB 1; Length 906;  
Best Local Similarity 100.0%; Pred. No. 3.6e-05;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3687 TAAGTACTAATATGATGTGATTTGAGTTT 3716  
|||||  
DB 47 TAAGTACTAATATGATGTGATTTGAGTTT 18

RESULT 3  
US-08-811-583-4/c  
; Sequence 4, Application US/08811583  
; Patent No. 6218142  
; GENERAL INFORMATION:  
; APPLICANT: Wassenecker, Michael  
; APPLICANT: Riedel, Leonhard  
; APPLICANT: Schiebel, Winfried  
; APPLICANT: Sanger, Heinz  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN  
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RDRP)  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & NEAVE  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/811,583  
; FILING DATE: 05-MAR-1997  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley, James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: MPG-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 25 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid

US-08-811-583-4  
; DESCRIPTION: /desc = "primer"

Query Match 0.7%; Score 25; DB 4; Length 25;  
Best Local Similarity 100.0%; Pred. No. 0.01;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1838 CCATCTGCTTCCAGATTCGTTATG 1862  
|||||  
DB 25 CCATCTGCTTCCAGATTCGTTATG 1

RESULT 4  
US-08-811-583-6/c  
; Sequence 6, Application US/08811583  
; Patent No. 6218142  
; GENERAL INFORMATION:  
; APPLICANT: Wassenecker, Michael  
; APPLICANT: Riedel, Leonhard  
; APPLICANT: Schiebel, Winfried  
; APPLICANT: Sanger, Heinz  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN  
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RDRP)  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & NEAVE  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/811,583  
; FILING DATE: 05-MAR-1997  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley, James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: MPG-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 25 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "primer"

US-08-811-583-6  
Query Match 0.7%; Score 25; DB 4; Length 25;  
Best Local Similarity 100.0%; Pred. No. 0.01;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2659 GCCGATGGATATCTCCAGCACCC 2683  
|||||  
DB 25 GCCGATGGATATCTCCAGCACCC 1

RESULT 5  
US-08-811-583-5/c  
; Sequence 5, Application US/08811583  
; Patent No. 6218142  
; GENERAL INFORMATION:  
; APPLICANT: Wassenecker, Michael





NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0356 US  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 54:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1750 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: HMC1NOT01  
CLONE: 9476  
US-09-234-613-54

Query Match 0.5%; Score 20; DB 3; Length 1750;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1916 TTGAGAAAGAGCATGCGAA 1935  
|||||  
DB 1355 TTGAGAAAGAGCATGCGAA 1374

RESULT 8  
US-08-811-583-7  
Sequence 7, Application US/08811583  
Patent No. 6218142  
GENERAL INFORMATION:  
APPLICANT: Wassenecker, Michael  
APPLICANT: Riedel, Leonhard  
APPLICANT: Schiebel, Winfried  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
POLYPEPTIDES HAVING THE ENZYMACTIC ACTIVITY OF AN  
RNA-DIRECTED RNA POLYMERASE (RdRP)  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/811,583  
FILING DATE: 05-MAR-1997  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley, James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: MPG-1  
TELEPHONE: 212-596-9000  
TELEFAX: 212-596-9090  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 34 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "primer"

US-08-811-583-7

Query Match 0.5%; Score 19; DB 4; Length 34;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 33 ACTCATCACTCCCTCAAG 51  
|||||  
DB 16 ACTCATCACTCCCTCAAG 34  
RESULT 9  
US-08-990-823-5/c  
Sequence 5, Application US/08990823D  
Patent No. 6228371  
GENERAL INFORMATION:  
APPLICANT: Nano, Francis  
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding  
Immunostimulatory Peptides  
FILE REFERENCE: 49086  
CURRENT APPLICATION NUMBER: US/08/990,823D  
CURRENT FILING DATE: 1997-12-15  
EARLIER APPLICATION NUMBER: US 96/10375  
EARLIER FILING DATE: 1996-06-14  
EARLIER APPLICATION NUMBER: 60/000,254  
EARLIER FILING DATE: 1995-06-15  
NUMBER OF SEQ ID NOS: 113  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 456  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
NAME/KEY: Modified base  
OTHER INFORMATION: n represents a or g or c or t/u  
US-08-990-823-5

Query Match 0.5%; Score 19; DB 4; Length 456;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2406 ACAGCAGATCCCAACAG 2424  
|||||  
DB 296 ACAGCAGATCCCAACAG 278

RESULT 10  
US-07-732-242C-8/c  
Sequence 8, Application US/07732242C  
Patent No. 5298399  
GENERAL INFORMATION:  
APPLICANT: Uozumi, Takeshi; Masaki, Haruhiko;  
APPLICANT: Hidaka, Makoto; Nakamura, Akira;  
APPLICANT: Maeda, Michihisa; Yoneta, Yasuo  
TITLE OF INVENTION: Gene of Urease  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Frisnauf, Holtz, Goodman & Woodward, P.C.  
STREET: 600 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10016-2088  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 0.72mb  
COMPUTER: IBM PC compatible (NEC PC-9801 RX)  
OPERATING SYSTEM: MS DOS  
SOFTWARE: ASCII Form  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/732,242C  
FILING DATE: 19910718  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JPN 2-210178  
FILING DATE: 10-AUG-1990

```

: ATTORNEY/AGENT INFORMATION:
: NAME: Goodman, Herbert
: REGISTRATION NUMBER: 17081
: REFERENCES/DOCKET NUMBER: 910532/HG
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212)972-1400
: TELEFAX: (212)370-1622
: TELEX: 236368
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6131 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: both
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-07-732-242C-8

```

Query Match 0.5%; Score 19; DB 1; Length 6131;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 19: Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 922 GACTGGTTTCACCTTTTC 940  
|||||  
Db 2524 GACTGGTTTCACCTTTTC 2506

```

RESULT 11
US-09-813-133A-3/C
; Sequence 3, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CLO01173
; CURRENT APPLICATION NUMBER: US/09/813, 133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 55827
; TYPE: DNA
; ORGANISM: Human
; US-09-813-133A-3

```

Qy 627 TTTCATATGAAATATATG 645  
|||  
Db 52586 TTTCATATGAAATATATG 52568

RESULT 12  
US-08-811-583-9/c  
Sequence 9, Application US/08811583  
Patent No. 6218142  
GENERAL INFORMATION:  
APPLICANT: Wassenecker, Michael  
APPLICANT: Riedel, Leonhard  
APPLICANT: Schiebel, Winfried  
APPLICANT: Sanger, Heinz  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN  
TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RdRp)  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York

COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/811,583  
FILING DATE: 05-MAR-1997  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley, James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: MPG-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-596-9000  
TELEFAX: 212-596-9090  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "primer"  
--S-08-811-583-9

Qy 3615 GACTACCTTTTGTCCTTA 3632  
 Db 35 GACTACCTTTTGTCCTTA 18

RESULT 13  
US-09-058-389A-20/c  
Sequence 20, Application US/09058389A  
Patent No. 6130065  
GENERAL INFORMATION:  
APPLICANT: Belt, Judith A.  
APPLICANT: Crawford, Charles R.  
APPLICANT: Patel, Divyen  
TITLE OF INVENTION: A NITROBENZYL MERCAPTOPYRINERIBOSIDE  
TITLE OF INVENTION: (NBMPR)-INSENSITIVE, EQUILIBRATIVE, NUCLEOSIDE TRANSPORT  
TITLE OF INVENTION: PROTEIN, NUCLEIC ACIDS ENCODING THE SAME AND METHODS OF  
TITLE OF INVENTION: USE  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David A. Jackson, Esq.  
STREET: 411 Hackensack Ave, Continental Plaza, 4th  
STREET: Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/058,389A  
FILING DATE: April 9, 1998  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 1340-1-013N  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800

```
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 461 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: other nucleic acid
;   DESCRIPTION: /desc = "intron 10(5')"
```

US-09-058-389A-20

Query Match 0.5%; Score 18; DB 3; Length 461;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2068 GGAAGCTGTAGTCAGCT 2085  
|||||  
DB 427 GGAAGCTGTAGTCAGCT 410

RESULT 14  
US-09-611-781-20/C  
; Sequence 20, Application US/09611781  
; Patent No. 6423829  
; GENERAL INFORMATION:  
; APPLICANT: Belt, Judith A.  
; APPLICANT: Crawford, Charles R.  
; APPLICANT: Patel, Divyen  
; TITLE OF INVENTION: A NITROBENZYL MERCAPTOPYRINERIBOSIDE  
; TITLE OF INVENTION: (NEMPR)-INSENSITIVE, EQUILIBRATIVE, NUCLEOSIDE TRANSPORT  
; TITLE OF INVENTION: PROTEIN, NUCLEIC ACIDS ENCODING THE SAME AND METHODS OF  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David A. Jackson, Esq.  
; STREET: 411 Hackensack Ave, Continental Plaza, 4th  
; STREET: Floor  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/611,781  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/058,389  
; FILING DATE: April 9, 1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 1340-1-013N  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 461 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "intron 10(5')"

US-09-611-781-20

Query Match 0.5%; Score 18; DB 4; Length 461;

Best Local Similarity 100.0%; Pred. No. 40;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2068 GGAAGCTGTAGTCAGCT 2085  
|||||  
DB 427 GGAAGCTGTAGTCAGCT 410

RESULT 15  
PCT-US95-08295-15  
; Sequence 15, Application PC/TUS9508295  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: BREAST SPECIFIC GENES AND PROTEINS  
; NUMBER OF SEQUENCES: 30  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/08295  
; FILING DATE: 30-JUN-1995  
; CLASSIFICATION:  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 481 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; PCT-US95-08295-15

Query Match 0.5%; Score 18; DB 5; Length 481;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2989 CAGGAAGCTGAAGGACAA 3006  
|||||  
DB 26 CAGGAAGCTGAAGGACAA 43

RESULT 16  
US-09-641-638-345/C  
; Sequence 345, Application US/09641638  
; Patent No. 6432648  
; GENERAL INFORMATION:  
; APPLICANT: Blumenfeld, Marta  
; APPLICANT: Bougueleret, Lydie  
; APPLICANT: Chumakov, Ilya  
; APPLICANT: Cohen, Annick  
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING  
; FILE REFERENCE: GENSET.051CP1  
; CURRENT APPLICATION NUMBER: US/09/641,638  
; CURRENT FILING DATE: 2000-08-16  
; PRIOR APPLICATION NUMBER: US 09/502,330  
; PRIOR FILING DATE: 2000-02-11  
; PRIOR APPLICATION NUMBER: US 60/133,200  
; PRIOR FILING DATE: 1999-05-07  
; PRIOR APPLICATION NUMBER: US 09/275,267  
; PRIOR FILING DATE: 1999-03-23  
; PRIOR APPLICATION NUMBER: US 60/119,917  
; PRIOR FILING DATE: 1999-02-12  
; NUMBER OF SEQ ID NOS: 1304  
; SOFTWARE: Patent.pm  
; SEQ ID NO 345  
; LENGTH: 1001  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
; FEATURE:  
; NAME/KEY: allele  
; LOCATION: 501

OTHER INFORMATION: 12-854-64 : polymorphic base A or G  
NAME/KEY: misc\_binding  
LOCATION: 481..500  
OTHER INFORMATION: 12-854-64.misl, potential  
NAME/KEY: misc\_binding  
LOCATION: 502..521  
OTHER INFORMATION: 12-854-64.mis2, potential complement  
NAME/KEY: primer\_bind  
LOCATION: 438..457  
OTHER INFORMATION: upstream amplification primer  
NAME/KEY: primer\_bind  
LOCATION: 942..962  
OTHER INFORMATION: downstream amplification primer, complement  
NAME/KEY: misc\_binding  
LOCATION: 489..513  
OTHER INFORMATION: 12-854-64 potential probe  
US-09-641-638-345

Query Match 0.5%; Score 18; DB 4; Length 1001;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3562 ATATATATATATGTAAT 3579  
|||||  
Db 182 ATATATATATGTAAT 165

## RESULT 17

US-09-134-001C-1922  
Sequence 1922, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 1922  
LENGTH: 1440  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-1922

Query Match 0.5%; Score 18; DB 4; Length 1440;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 661 TCGTCCATATGTCAGAAA 678  
|||||  
Db 603 TCGTCCATATGTCAGAAA 620

## RESULT 18

US-08-383-750-3/C  
Sequence 3, Application US/08383750  
Patent No. 5744301  
GENERAL INFORMATION:  
APPLICANT: Birkenbach, Mark  
APPLICANT: Kieff, Elliott  
TITLE OF INVENTION: Epstein Barr Virus Induced Genes  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue, N.W.,  
CITY: Washington  
STATE: D.C.

COUNTRY: U.S.A.  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/383,750  
FILING DATE: Herewith  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: Fox, Samuel, L.  
REGISTRATION NUMBER: 30,353  
REFERENCE/DOCKET NUMBER: 0827.3300001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1643 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 34..1116  
US-08-383-750-3

Query Match 0.5%; Score 18; DB 1; Length 1643;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1382 GTTGATGAGAGTGGGAG 1399  
|||||  
Db 540 GTTGATGAGAGTGGGAG 523

## RESULT 19

US-08-352-678-3/C  
Sequence 3, Application US/08352678  
Patent No. 6043351  
GENERAL INFORMATION:  
APPLICANT: Birkenbach, Mark  
APPLICANT: Kieff, Elliott  
TITLE OF INVENTION: EPSTEIN BARR VIRUS INDUCED GENES  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/352,678  
FILING DATE: 30-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/980,518  
FILING DATE: 25-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Gates, Edward R.  
REGISTRATION NUMBER: 31,616  
REFERENCE/DOCKET NUMBER: B0801/7044  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1643 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 34...1116  
US-08-352-678-3

Query Match 0.5%; Score 18; DB 3; Length 1643;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1382 GTTGATGAGGAGTGGAG 1399  
|||||  
DB 540 GTTGATGAGGAGTGGAG 523

RESULT 20  
PCT-US93-09636-3/c  
Sequence 3, Application PC/TUS9309636  
GENERAL INFORMATION:  
APPLICANT: Birkenbach, Mark  
TITLE OF INVENTION: Epstein Barr Virus Induced Genes  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue N.W., Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3934

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/09636  
FILING DATE: herewith  
CLASSIFICATION:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1643 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 34...1116  
PCT-US93-09636-3

Query Match 0.5%; Score 18; DB 5; Length 1643;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1382 GTTGATGAGGAGTGGAG 1399  
|||||  
DB 540 GTTGATGAGGAGTGGAG 523

RESULT 21  
US-08-454-720A-35  
Sequence 35, Application US/08454720A  
Patent No. 5766888  
GENERAL INFORMATION:

APPLICANT: Sobol, Robert E.  
APPLICANT: Green, Mark R.  
APPLICANT: Kawasaki, Ernest S.  
TITLE OF INVENTION: Detection of Carcinoma Metastases by  
TITLE OF INVENTION: Nucleic Acid Amplification  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffman-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.01, version  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/454,720A  
FILING DATE: May 31, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/357,565  
FILING DATE: December 16, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Petry, Douglas  
REGISTRATION NUMBER: 35321  
REFERENCE/DOCKET NUMBER: 9178  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 814-2974  
TELEFAX: (510) 814-2977  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1990 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-454-720A-35

Query Match 0.5%; Score 18; DB 1; Length 1990;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 576 AGAAATTTTCTTTT 593  
|||||  
DB 1445 AGAAATTTTCTTTT 1462

RESULT 22  
US-09-323-872A-45/c  
Sequence 45, Application US/09323872A  
Patent No. 6395539  
GENERAL INFORMATION:  
APPLICANT: Coschigano, Peter  
TITLE OF INVENTION: Compositions and Methods for Bioremediation  
FILE REFERENCE: OHU-03640  
CURRENT APPLICATION NUMBER: US/09/323,872A  
CURRENT FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: 09/072,433  
PRIOR FILING DATE: 1998-05-04  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 45  
LENGTH: 2038  
TYPE: DNA  
ORGANISM: Thauera aromatica  
US-09-323-872A-45

Query Match 0.5%; Score 18; DB 4; Length 2038;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1368 TTGCTGTTCTCTTTGTTG 1385  
|||||  
Db 1212 TTGCTGTTCTCTTTGTTG 1195

## RESULT 23

US-09-620-412C-328/C  
; Sequence 328, Application US/09620412C  
; Patent No. 6448234

## ; GENERAL INFORMATION:

; APPLICANT: Steven P. Fling  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; FILE REFERENCE: 210121.469C7

; CURRENT APPLICATION NUMBER: US/09/620,412C

; CURRENT FILING DATE: 2000-07-20

; NUMBER OF SEQ ID NOS: 363

; SOFTWARE: FastSeq for Windows Version 3.0/4.0

; SEQ ID NO 328

; LENGTH: 2148

; TYPE: DNA

; ORGANISM: Chlamydia trachomatis

US-09-620-412C-328

Query Match 0.5%; Score 18; DB 4; Length 2148;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3657 ATGCAGTACTTAAGAAA 3674

|||||

Db 1700 ATGCAGTACTTAAGAAA 1683

## RESULT 24

US-08-118-101A-3

; Sequence 3, Application US/08118101A

; Patent No. 5620892

## ; GENERAL INFORMATION:

; APPLICANT: Kutz, Stephen E.

; APPLICANT: Knickerbocker, Aron M.

; APPLICANT: McCullough, John R.

; TITLE OF INVENTION: A STRAIN OF SACCCHAROMYCES CEREVISIAE

; TITLE OF INVENTION: EXPRESSING THE GENE ENCODING POTASSIUM TRANSPORTER MINK

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Burton Rodney

; STREET: P.O. Box 4000

; CITY: Princeton

; STATE: New Jersey

; COUNTRY: U.S.A.

; ZIP: 08543-4000

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/118,101A

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Gaul, Timothy J.

; REGISTRATION NUMBER: 33,111

; REFERENCE/DOCKET NUMBER: DC27

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (609) 252-5901

; TELEFAX: (609) 252-4526

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2669 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

;

;

;

;

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..2669

US-08-118-101A-3

Query Match 0.5%; Score 18; DB 1; Length 2669;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1376 TCTTTTGTGTGAGGAG 1393

|||||

Db 1147 TCTTTTGTGTGAGGAG 1164

## RESULT 25

US-08-348-143-3

; Sequence 3, Application US/08348143

; Patent No. 5506205

## ; GENERAL INFORMATION:

; APPLICANT: TAJIMA, HISAO

; APPLICANT: KITAGAWA, KOICHIRO

; APPLICANT: OHNO, HIROYUKI

; APPLICANT: UENO, TOSHIO

; TITLE OF INVENTION: A No. 5506205el Polypeptide of Protein p140 and DNAs

; TITLE OF INVENTION: encoding it

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS

; STREET: 2100 Pennsylvania Avenue, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20037-3202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/348,143

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 315806/1993

; FILING DATE: 24-NOV-1993

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)293-7060

; TELEFAX: (202)293-7860

; TELEX: 6491103

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4027 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

; ORIGINAL SOURCE:

; ORGANISM: rat

; TISSUE TYPE: skeletal muscle myoblast

; CELL LINE: L6

US-08-348-143-3

Query Match 0.5%; Score 18; DB 1; Length 4027;

Best Local Similarity 100.0%; Pred. No. 44;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ATTGGTTGTGCTTCAGT 91

|||||

Db 3523 ATTGGTTGTGCTTCAGT 3540

```

RESULT 26
US-08-348-143-4
; Sequence 4, Application US/08348143
; Patent No. 5506205
; GENERAL INFORMATION:
; APPLICANT: TAJIMA, HISAO
; APPLICANT: KITAGAWA, KOICHIRO
; APPLICANT: OHNO, HIROYUKI
; APPLICANT: UENO, TOSHIO
; TITLE OF INVENTION: A No. 5506205el Polypeptide of Protein p140 and DNAs
; TITLE OF INVENTION: encoding it
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/348,143
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 315806/1993
; FILING DATE: 24-NOV-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4027 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: rat
; TISSUE TYPE: skeletal muscle myoblast
; CELL LINE: L6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 262..3243
; IDENTIFICATION METHOD: by similarity to some other pattern
US-08-348-143-4
Query Match 0.5%; Score 18; DB 1; Length 4027;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ATTGTGTTGCTTCAGT 91
|||||
Db 3523 ATTGTGTTGCTTCAGT 3540
|||||

RESULT 27
US-08-571-785-3
; Sequence 3, Application US/08571785
; Patent No. 5804411
; GENERAL INFORMATION:
; APPLICANT: TAJIMA, HISAO
; APPLICANT: KITAGAWA, KOICHIRO
; APPLICANT: OHNO, HIROYUKI
; APPLICANT: UENO, TOSHIO
; TITLE OF INVENTION: A No. 5804411el Polypeptide of Protein p140
; TITLE OF INVENTION: and DNAs encoding it
; NUMBER OF SEQUENCES: 16

```

```
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/348,143
; FILING DATE: 23-NOV-1994
; APPLICATION NUMBER: JP 315806/1993
; FILING DATE: 24-NOV-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4027 base pairs
; TYPE: nucleic acid
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: rat
; TISSUE TYPE: skeletal muscle myoblast
; CELL LINE: L6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 262..3243
; IDENTIFICATION METHOD: by similarity to some other pattern
US-08-571-785-4
```

Query Match 0.5%; Score 18; DB 1; Length 4027;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 74 ATTGGTGTGCTTCAGT 91
|||||
DB 3523 ATTGGTGTGCTTCAGT 3540
```

## RESULT 29

```
US-09-192-435-3
; Sequence 3, Application US/09192435
; Patent No. 6303320
; GENERAL INFORMATION:
; APPLICANT: TAJIMA, HISAO
; APPLICANT: KITAGAWA, KOICHIRO
; APPLICANT: OHNO, HIROYUKI
; APPLICANT: UENO, TOSHIO
; TITLE OF INVENTION: A No. 6303320el Polypeptide of Protein p140
; TITLE OF INVENTION: and DNAs encoding it
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/192,435
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/571,785
; FILING DATE: 13-DEC-1995
; APPLICATION NUMBER: 08/348,143
; FILING DATE: 23-NOV-1994
; APPLICATION NUMBER: JP 315806/1993
; FILING DATE: 24-NOV-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4027 base pairs
; TYPE: nucleic acid
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: rat
```

```
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4027 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: rat
; TISSUE TYPE: skeletal muscle myoblast
; CELL LINE: L6
US-09-192-435-3
```

Query Match 0.5%; Score 18; DB 4; Length 4027;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 74 ATTGGTGTGCTTCAGT 91
|||||
DB 3523 ATTGGTGTGCTTCAGT 3540
```

## RESULT 30

```
US-09-192-435-4
; Sequence 4, Application US/09192435
; Patent No. 6303320
; GENERAL INFORMATION:
; APPLICANT: TAJIMA, HISAO
; APPLICANT: KITAGAWA, KOICHIRO
; APPLICANT: OHNO, HIROYUKI
; APPLICANT: UENO, TOSHIO
; TITLE OF INVENTION: A No. 6303320el Polypeptide of Protein p140
; TITLE OF INVENTION: and DNAs encoding it
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/192,435
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/571,785
; FILING DATE: 13-DEC-1995
; APPLICATION NUMBER: 08/348,143
; FILING DATE: 23-NOV-1994
; APPLICATION NUMBER: JP 315806/1993
; FILING DATE: 24-NOV-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4027 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: rat
```



TISSUE TYPE: skeletal muscle myoblast  
CELL LINE: L6  
NAME/KEY: CDS  
LOCATION: 262..3243  
IDENTIFICATION METHOD: by similarity to some other pattern  
US-09-192-435-4

Query Match 0.5%; Score 18; DB 4; Length 4027;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ATTGGTTGCTTCAGT 91  
|||||  
DB 3523 ATTGGTTGCTTCAGT 3540

## RESULT 31

US-09-558-340-3  
Sequence 3, Application US/09558340  
Patent No. 6432913  
GENERAL INFORMATION:  
APPLICANT: TAJIMA, HISAO  
APPLICANT: KITAGAWA, KOICHIRO  
APPLICANT: OHNO, HIROYUKI  
APPLICANT: UENO, TOSHIO  
TITLE OF INVENTION: A No. 6432913el Polypeptide of Protein p140  
TITLE OF INVENTION: and DNAs encoding it  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS  
STREET: 2100 Pennsylvania Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20037-3202  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/558,340  
FILING DATE: 26-APR-2000  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/192,435  
FILING DATE: 08-JAN-1998  
APPLICATION NUMBER: 08/571,785  
FILING DATE: 13-DEC-1995  
APPLICATION NUMBER: 08/348,143  
FILING DATE: 23-NOV-1994  
APPLICATION NUMBER: JP 315806/1993  
FILING DATE: 24-NOV-1993  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)293-7060  
TELEFAX: (202)293-7860  
TELEX: 6491103  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4027 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ORIGINAL SOURCE:  
ORGANISM: rat  
TISSUE TYPE: skeletal muscle myoblast  
CELL LINE: L6

Query Match 0.5%; Score 18; DB 4; Length 4027;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ATTGGTTGCTTCAGT 91  
|||||  
DB 3523 ATTGGTTGCTTCAGT 3540

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 74 ATTGGTTGCTTCAGT 91  
|||||  
DB 3523 ATTGGTTGCTTCAGT 3540

## RESULT 32

US-09-558-340-4  
Sequence 4, Application US/09558340  
Patent No. 6432913  
GENERAL INFORMATION:  
APPLICANT: TAJIMA, HISAO  
APPLICANT: KITAGAWA, KOICHIRO  
APPLICANT: OHNO, HIROYUKI  
APPLICANT: UENO, TOSHIO  
TITLE OF INVENTION: A No. 6432913el Polypeptide of Protein p140  
TITLE OF INVENTION: and DNAs encoding it  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS  
STREET: 2100 Pennsylvania Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20037-3202  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/558,340  
FILING DATE: 26-APR-2000  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/192,435  
FILING DATE: 08-JAN-1998  
APPLICATION NUMBER: 08/571,785  
FILING DATE: 13-DEC-1995  
APPLICATION NUMBER: 08/348,143  
FILING DATE: 23-NOV-1994  
APPLICATION NUMBER: JP 315806/1993  
FILING DATE: 24-NOV-1993  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)293-7060  
TELEFAX: (202)293-7860  
TELEX: 6491103  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4027 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ORIGINAL SOURCE:  
ORGANISM: rat  
TISSUE TYPE: skeletal muscle myoblast  
CELL LINE: L6  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 262..3243  
IDENTIFICATION METHOD: by similarity to some other pattern  
US-09-558-340-4

Query Match 0.5%; Score 18; DB 4; Length 4027;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ATTGGTTGCTTCAGT 91  
|||||  
DB 3523 ATTGGTTGCTTCAGT 3540



```
; APPLICANT: Crawford, Charles R.
; APPLICANT: Patel, Divyen
; TITLE OF INVENTION: A NITROBENZYL MERCAPTOPYRIMIDINOSIDE
; TITLE OF INVENTION: (NMPR)-INSENSITIVE, EQUILIBRATIVE, NUCLEOSIDE TRANSPORT
; TITLE OF INVENTION: PROTEIN, NUCLEIC ACIDS ENCODING THE SAME AND METHODS OF
; TITLE OF INVENTION: USE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/611,781
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/058,389
; FILING DATE: April 9, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1340-1-013N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6354 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; US-09-611-781-5

Query Match 0.5%; Score 18; DB 4; Length 6354;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2068 GGAAGCTGTAGATCAGCT 2085
Db 6320 GGAAGCTGTAGATCAGCT 6303
|||||

RESULT 37
US-09-227-219-130/C
; Sequence 130, Application US/08927219
; Patent No. 6187533
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme I.
; APPLICANT: Yamagata, Kazuya
; APPLICANT: Oda, Naohisa
; APPLICANT: Katsaki, Pamela J.
; APPLICANT: Furuta, Hiroto
; APPLICANT: Horikawa, Yukio
; APPLICANT: Menzel, Stephen
; TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY
; TITLE OF INVENTION: GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA
; TITLE OF INVENTION: AND HNF-4ALPHA
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
```

```
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,219
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,679
; FILING DATE: 30-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,056
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/025,719
; FILING DATE: 10-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: ARCD:272
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 130:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10014 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-927-219-130

Query Match 0.5%; Score 18; DB 4; Length 10014;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1384 TGATGAGGAGTGGGAGAA 1401
Db 4194 TGATGAGGAGTGGGAGAA 4177
|||||

RESULT 38
US-08-991-840A-1
; Sequence 1, Application US/08991840A
; Patent No. 6261570
; GENERAL INFORMATION:
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Bruce Crise
; APPLICANT: Mark Steve Oberste
; APPLICANT: Shannon Schmura
; TITLE OF INVENTION: Live Attenuated Virus Vaccines for Eastern Equine Encephali
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Charles H. Harris
; STREET: USA MRMC - MRMC-JA
; CITY: FORT DETRICK, FREDERICK
; STATE: MARYLAND
; COUNTRY: USA
; ZIP: 21702-5012
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.5
; SOFTWARE: Microsoft Word 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,840A
; FILING DATE: December 16, 1997
; CLASSIFICATION: 424
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: Provisional Application 60/047162,
; FILING DATE: May 20, 1997
; APPLICATION NUMBER: Provisional Application 60/053,652
; FILING DATE: July 24, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles H. Harris
; REGISTRATION NUMBER: 34,616
; REFERENCE/DOCKET NUMBER: 003/058/SAP RIID 96-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 619-2065
; TELEFAX: (301) 619-5034
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11492 base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Linear
; FEATURE: OTHER INFORMATION: N at all occurrences is = unknown.
; FEATURE: OTHER INFORMATION: K at all
; FEATURE: occurrences is = G or T
US-08-991-840A-1

Query Match      0.5%; Score 18; DB 4; Length 11492;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 GTGCTGACAAATCATCA 380
|||||
Db 2636 GTGCTGACAAATCATCA 2653

RESULT 39
US-09-819-993-3/c
; Sequence 3, Application US/09819993
; Patent No. 6436692
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; MOLECULE TYPE: DNA
; FILE REFERENCE: CL001195
; CURRENT APPLICATION NUMBER: US/09/819,993
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 28001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(28001)
; OTHER INFORMATION: n = A,T,C or G
US-09-819-993-3

Query Match      0.5%; Score 18; DB 4; Length 28001;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3409 TTTCATTAGCTTCCCTG 3426
|||||
Db 19180 TTTCATTAGCTTCCCTG 19163

RESULT 40
US-09-215-694-19/c
; Sequence 19, Application US/09215694B
; Patent No. 6391583
; GENERAL INFORMATION:
; APPLICANT: Wisconsin Alumni Research Foundation
; APPLICANT: Hutchinson, Charles R.
; APPLICANT: Kennedy, Jonathan n.m.i.
```

```
; APPLICANT: Park, Cheonseok n.m.i
; TITLE OF INVENTION: METHOD OF PRODUCING ANTIHYPERCHOLESTEROLEMIC AGENTS
; FILE REFERENCE: 960296.95718
; CURRENT APPLICATION NUMBER: US/09/215,694B
; CURRENT FILING DATE: 1999-12-18
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 19
; LENGTH: 31328
; TYPE: DNA
; ORGANISM: Aspergillus terreus
US-09-215-694-19

Query Match      0.5%; Score 18; DB 4; Length 31328;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2048 GTTCTGACAGAGCAAA 2065
|||||
Db 27833 GTTCTGACAGAGCAAA 27816

RESULT 41
US-09-453-702B-116/c
; Sequence 116, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 116:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45175
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 116:
US-09-453-702B-116

Query Match      0.5%; Score 18; DB 4; Length 45175;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 2270 AGAATATTATTCACAAAT 2287  
|||||  
Db 56 AGAATATTATTCACAAAT 39

## RESULT 42

US-09-188-930-206  
; Sequence 206, Application US/09188930A  
; Patent No. 6150502  
; GENERAL INFORMATION:

; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James Greg  
; TITLE OF INVENTION: Compositions Isolated From Skin Cells  
; FILE REFERENCE: 11000.1011c1  
; CURRENT APPLICATION NUMBER: US/09/188,930A  
; CURRENT FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 348  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 206  
; LENGTH: 480  
; TYPE: DNA  
; ORGANISM: Mouse  
US-09-188-930-206

Query Match 0.5%; Score 17; DB 3; Length 480;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 TGCTTCAGTTTCAGTCA 99  
|||||  
Db 262 TGCTTCAGTTTCAGTCA 278

## RESULT 43

US-08-460-739-2  
; Sequence 2, Application US/08460739  
; Patent No. 5698416  
; GENERAL INFORMATION:

; APPLICANT: Wolf, Marcia K.  
; APPLICANT: Cassels, Frederick J.  
; APPLICANT: Bell, Brian A.  
; TITLE OF INVENTION: Improved Methods for Production of  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Glenna Hendricks  
; STREET: 9669 A Main Street  
; CITY: Fairfax  
; STATE: VA  
; COUNTRY: US  
; ZIP: 22031

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/460,739  
; FILING DATE: 20-APR-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hendricks, Glenna M.  
; REGISTRATION NUMBER: 32,535  
; REFERENCE/DOCKET NUMBER: wolf2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 425-4250  
; TELEFAX: (703) 425-2767  
; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 512 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: DNA for encoding CFAL protein  
US-08-460-739-2

Query Match 0.5%; Score 17; DB 1; Length 512;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3212 GCATCAAAAACCTTTGA 3228  
|||||  
Db 174 GCATCAAAAACCTTTGA 190

## RESULT 44

US-09-134-001C-1886/c  
; Sequence 1886, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 1886  
; LENGTH: 777  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-1886

Query Match 0.5%; Score 17; DB 4; Length 777;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 924 CTGGTTTCACCTTTTC 940  
|||||  
Db 439 CTGGTTTCACCTTTTC 423

## RESULT 45

US-08-961-527-264/c  
; Sequence 264, Application US/08961527  
; Patent No. 6420135  
; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 264:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 795 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-264

Query Match 0.5%; Score 17; DB 4; Length 795;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0;

QY 2865 ACTTCCAAAGACTGGT 2881  
|||||  
DB 655 ACTTCCAAAGACTGGT 639

## RESULT 46

US-09-166-350-10/c  
Sequence 10, Application US/09166350A  
Patent No. 6440663  
GENERAL INFORMATION:  
APPLICANT: Scanlan, Matthew  
APPLICANT: Chen, Yao  
APPLICANT: Stockert, Elisabeth  
APPLICANT: Old, Lloyd  
APPLICANT: Jager, Elke  
APPLICANT: Knuth, Alex  
TITLE OF INVENTION: Renal Cancer Associated Antigens and  
TITLE OF INVENTION: Uses Therefor  
FILE REFERENCE: L0461/7051  
CURRENT APPLICATION NUMBER: US/09/166,350A  
EARLIER FILING DATE: 1998-10-05  
EARLIER APPLICATION NUMBER: US 09/166,350  
EARLIER FILING DATE: 1998-10-05  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 10  
LENGTH: 833  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-166-350-10

Query Match 0.5%; Score 17; DB 4; Length 833;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0;

QY 742 CTGTTATAGCTCTTTA 758  
|||||  
DB 407 CTGTTATAGCTCTTTA 391

## RESULT 47

US-09-188-930-36  
Sequence 36, Application US/09188930A  
Patent No. 6150502  
GENERAL INFORMATION:  
APPLICANT: Watson, James D.  
APPLICANT: Strachan, Lorna  
APPLICANT: Sleeman, Matthew

APPLICANT: Onrust, Rene  
APPLICANT: Murison, James Greg  
TITLE OF INVENTION: Compositions Isolated From Skin Cells  
TITLE OF INVENTION: and Methods For Their Use  
FILE REFERENCE: 11000.1011c1  
CURRENT APPLICATION NUMBER: US/09/188,930A  
CURRENT FILING DATE: 1998-11-09  
NUMBER OF SEQ ID NOS: 348  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 36  
LENGTH: 896  
TYPE: DNA  
ORGANISM: mouse  
US-09-188-930-36

Query Match 0.5%; Score 17; DB 3; Length 896;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0;

QY 83 TGCTTCAGTTTCAGTCA 99  
|||||  
DB 194 TGCTTCAGTTTCAGTCA 210

## RESULT 48

US-08-671-525B-7  
Sequence 7, Application US/08671525B  
Patent No. 5703220  
GENERAL INFORMATION:  
APPLICANT: Yamada, Tadataka  
APPLICANT: Gantz, Ira  
TITLE OF INVENTION: Genes Encoding Melanocortin Receptors  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.  
STREET: P.O. Box 828  
CITY: Bloomfield Hills  
STATE: MI  
COUNTRY: US  
ZIP: 48303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/671,525B  
FILING DATE: June 27, 1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, Deann F.  
REGISTRATION NUMBER: 36683  
REFERENCE/DOCKET NUMBER: 2115-000853DVB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (810)641-1600  
TELEFAX: (810)641-0270  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 996 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..996  
US-08-671-525B-7

Query Match 0.5%; Score 17; DB 1; Length 996;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2850 TTTCATTCAGTGGAC 2866  
|||||  
Db 422 TTTCATTCAGTGGAC 438

RESULT 49  
US-08-672-109B-7  
; Sequence 7, Application US/08672109B  
; Patent No. 5710265  
; GENERAL INFORMATION:  
; APPLICANT: Yamada, Tadataka  
; APPLICANT: Gantz, Ira  
; TITLE OF INVENTION: Genes Encoding Melanocortin Receptors  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.  
; STREET: P.O. Box 828  
; CITY: Bloomfield Hills  
; STATE: MI  
; COUNTRY: US  
; ZIP: 48303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/672,109B  
; FILING DATE: June 27, 1996  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, Deann F.  
; REGISTRATION NUMBER: 36683  
; REFERENCE/DOCKET NUMBER: 2115-000853DVC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (810)641-1600  
; TELEFAX: (810)641-0270  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 996 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..996  
US-08-672-109B-7

Query Match 0.5%; Score 17; DB 1; Length 996;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2850 TTTCATTCAGTGGAC 2866  
|||||  
Db 422 TTTCATTCAGTGGAC 438

RESULT 50  
US-08-842-045-7  
; Sequence 7, Application US/08842045  
; Patent No. 5817787  
; GENERAL INFORMATION:  
; APPLICANT: Yamada, Tadataka  
; APPLICANT: Gantz, Ira  
; TITLE OF INVENTION: Genes Encoding Melanocortin Receptors

; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.  
; STREET: P.O. Box 828  
; CITY: Bloomfield Hills  
; STATE: MI  
; COUNTRY: US  
; ZIP: 48303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/842,045  
; FILING DATE:  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, Deann F.  
; REGISTRATION NUMBER: 36683  
; REFERENCE/DOCKET NUMBER: 2115-000853DVE  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (810)641-1600  
; TELEFAX: (810)641-0270  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 996 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..996  
US-08-842-045-7

Query Match 0.5%; Score 17; DB 1; Length 996;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2850 TTTCATTCAGTGGAC 2866  
|||||  
Db 422 TTTCATTCAGTGGAC 438

Search completed: November 6, 2002, 03:36:45  
Job time : 687 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 23:31:02 ; Search time 158 Seconds

(without alignments)  
7867.093 Million cell updates/sec

Title: US-09-782-874-1

Perfect score: 3731

Sequence: 1 GAATAATCTTTACTTACTT.....AGTTCACTTCTCTCTAAA 3731

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 310279 seqs, 166577418 residues

Word size : 0

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database :

Published\_Applications\_NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US05\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3731	100.0	3731	10	US-09-782-874-1
2	25	0.7	25	10	US-09-782-874-4
3	25	0.7	25	10	US-09-782-874-6
4	24	0.6	25	10	US-09-782-874-5
5	20	0.5	398	10	US-09-998-598-194
6	20	0.5	407	10	US-09-962-436-82
7	20	0.5	409	10	US-09-998-598-772
8	20	0.5	413	10	US-09-998-598-2240
9	20	0.5	1750	10	US-09-840-787-54
10	19	0.5	34	10	US-09-782-874-7
11	19	0.5	267	10	US-09-563-817-946
12	19	0.5	461	10	US-09-864-761-11608
13	19	0.5	4656	10	US-09-764-877-2690
14	18	0.5	35	10	US-09-782-874-9
15	18	0.5	256	10	US-09-923-876-1838
16	18	0.5	358	10	US-09-960-352-1347
17	18	0.5	385	10	US-09-867-701-4817
18	18	0.5	474	10	US-09-864-761-11283
19	18	0.5	500	10	US-09-783-590-8123

c 20	18	0.5	1016	10	US-09-925-239-176	Sequence 176, App
c 21	18	0.5	1467	10	US-09-815-242-7347	Sequence 7347, App
c 22	18	0.5	1497	10	US-09-887-576-629	Sequence 629, App
c 23	18	0.5	1944	10	US-09-887-576-619	Sequence 619, App
c 24	18	0.5	2148	10	US-09-841-132-328	Sequence 328, App
c 25	18	0.5	3048	12	US-10-007-693-80	Sequence 80, Appl
c 26	18	0.5	3760	10	US-09-822-849A-162	Sequence 162, App
c 27	18	0.5	3781	12	US-10-052-586-453	Sequence 453, App
c 28	18	0.5	4348	10	US-09-879-445-1	Sequence 1, Appl
c 29	18	0.5	32192	10	US-09-764-877-3657	Sequence 3657, App
c 30	18	0.5	49744	10	US-09-927-091-4	Sequence 4, Appl
c 31	18	0.5	170834	10	US-09-835-232-7	Sequence 7, Appl
c 32	17	0.5	150	10	US-09-923-876-3754	Sequence 754, App
c 33	17	0.5	213	10	US-09-878-574-9084	Sequence 9084, App
c 34	17	0.5	258	10	US-09-878-574-7593	Sequence 7593, App
c 35	17	0.5	269	10	US-09-969-373-41	Sequence 41, Appl
c 36	17	0.5	274	10	US-09-294-093B-1096	Sequence 1096, App
c 37	17	0.5	308	10	US-09-864-761-31520	Sequence 31520, App
c 38	17	0.5	329	10	US-09-880-107-706	Sequence 706, App
c 39	17	0.5	353	10	US-09-917-800A-623	Sequence 623, App
c 40	17	0.5	368	10	US-09-864-761-13956	Sequence 13956, App
c 41	17	0.5	387	10	US-09-974-300-7289	Sequence 7289, App
c 42	17	0.5	391	10	US-09-864-761-30520	Sequence 30520, App
c 43	17	0.5	391	10	US-09-563-817-878	Sequence 878, App
c 44	17	0.5	406	10	US-09-983-965-401	Sequence 401, App
c 45	17	0.5	411	10	US-09-867-550-1095	Sequence 1095, App
c 46	17	0.5	419	10	US-09-960-352-10698	Sequence 10698, App
c 47	17	0.5	421	10	US-09-880-107-2705	Sequence 2705, App
c 48	17	0.5	430	10	US-09-864-761-14986	Sequence 14986, App
c 49	17	0.5	436	10	US-09-864-761-2826	Sequence 2826, App
c 50	17	0.5	436	10	US-09-983-965-2934	Sequence 2934, App

#### ALIGNMENTS

#### RESULT 1

US-09-782-874-1

; Sequence 1, Application US/09782874

; Patent No. US20010023067A1

; GENERAL INFORMATION:

; APPLICANT: Wassenecker, Michael

; Riedel, Leonhard

; Schiebel, Winfried

; Sanger, Heinz

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING

; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN

; RNA-DIRECTED RNA POLYMERASE (RDRP)

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FISH & NEAVE

; STREET: 1251 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10020

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC Compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/782.874

; FILING DATE: 08-Feb-2001

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/811,583

; FILING DATE: 05-MAR-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Haley, James F.

; REGISTRATION NUMBER: 27,794

; REFERENCE/DOCKET NUMBER: MPG-1

; TELECOMMUNICATION INFORMATION:



TELEPHONE: 212-596-9000  
TELEFAX: 212-596-9090  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3731 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Tomato  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 194...3535  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
us-09-782-874-1

Query Match 100.0%; Score 3731; DB 10; Length 3731;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3731; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAATATCTTTACTTACCTTACCAGGGATTGACTCATCTACCTACTACTGCTGGGTAGTTTAT 60  
DB 1 GAAATATCTTTACTTACCTTACCAGGGATTGACTCATCTACCTACTACTGCTGGGTAGTTTAT 60  
QY 61 GTTGCGATATAAATTTGGTGTGCTTCAAGTTTCACTACTACTACTGCTGGGTAGTTTAT 120  
DB 61 GTTGCGATATAAATTTGGTGTGCTTCAAGTTTCACTACTACTACTGCTGGGTAGTTTAT 120  
QY 121 TTTGCGATACCTCAGGGGTATTCAGTTGGTGTAGCTATTCAGATTTGAAGTGCAGACTG 180  
DB 121 TTTGCGATACCTCAGGGGTATTCAGTTGGTGTAGCTATTCAGATTTGAAGTGCAGACTG 180  
QY 181 GGAATTTGGCTACATGGGAAGCAATTCAGTTTTCGGATTCCTTACTCTCTCTGC 240  
DB 181 GGAATTTGGCTACATGGGAAGCAATTCAGTTTTCGGATTCCTTACTCTCTCTGC 240  
QY 241 GGAAGTGGTTAAGTCAATCTTAGAGAAATATACAGGATATGGAATGTAATGCAATGGA 300  
DB 241 GGAAGTGGTTAAGTCAATCTTAGAGAAATATACAGGATATGGAATGTAATGCAATGGA 300  
QY 301 GGTAAACAGTCAAGAGGAGTCTAGAGCAATTCGCAAGTTCATTTCCGAGCAACAT 360  
DB 301 GGTAAACAGTCAAGAGGAGTCTAGAGCAATTCGCAAGTTCATTTCCGAGCAACAT 360  
QY 361 AAGTGTGCAAAATCATCACTTTGGCTAATACAGGCTGTATTTGGCTCTCTTATTT 420  
DB 361 AAGTGTGCAAAATCATCACTTTGGCTAATACAGGCTGTATTTGGCTCTCTTATTT 420  
QY 421 GAAGGCTGGGAATCAAACTGATATTGCAACTGCGGGCATATGTGGATCAGATGGA 480  
DB 421 GAAGGCTGGGAATCAAACTGATATTGCAACTGCGGGCATATGTGGATCAGATGGA 480  
QY 481 TGCCATAAATTTGAATTTCCGATGTCAGATATCAGATGACAAGTTGCAAGTTGGGAAG 540  
DB 481 TGCCATAAATTTGAATTTCCGATGTCAGATATCAGATGACAAGTTGCAAGTTGGGAAG 540  
QY 541 TACAGAGTTTCAATTCATTTGGCATGGATGGAATTTTCTTTTATCTAG 600  
DB 541 TACAGAGTTTCAATTCATTTGGCATGGATGGAATTTTCTTTTATCTAG 600  
QY 601 TGGTTCAGTGTACTATAAACTTCAGTTTCATATGAATATATGCGAGTTGCTCCA 660  
DB 601 TGGTTCAGTGTACTATAAACTTCAGTTTCATATGAATATATGCGAGTTGCTCCA 660  
QY 661 TCGTCCATATGGTCAAAATGCTCAGTTTCTCCTCATACAGTTATTTGGTGCCTCGGAT 720  
DB 661 TCGTCCATATGGTCAAAATGCTCAGTTTCTCCTCATACAGTTATTTGGTGCCTCGGAT 720  
QY 721 CTATAAGAGACTTGAAACTCTGTTATAGCTTCTTAAAGAACTCCTGATGATCAGTG 780  
DB 721 CTATAAGAGACTTGAAACTCTGTTATAGCTTCTTAAAGAACTCCTGATGATCAGTG 780

QY 781 GGTGAGGACAACAGATTTCCCTCCATCTTGGATAGGCTATCTTCTAGCTTATGTTTGA 840  
DB 781 GGTGAGGACAACAGATTTCCCTCCATCTTGGATAGGCTATCTTCTAGCTTATGTTTGA 840  
QY 841 GTTCCGTAGGGGTGTTGCTCTTCCAAATTTCCGAGGAAAGTTTTTCCACTATGAGAACG 900  
DB 841 GTTCCGTAGGGGTGTTGCTCTTCCAAATTTCCGAGGAAAGTTTTTCCACTATGAGAACG 900  
QY 901 TGAAAACAATATTACTTTACAGACTGGTTTCCACCTTTTTCGCTCTCTCAAAATCGGCTCT 960  
DB 901 TGAAAACAATATTACTTTACAGACTGGTTTCCACCTTTTTCGCTCTCTCAAAATCGGCTCT 960  
QY 961 GGTTCCTCAATGTCCAGCTCCGGAAGGAATTTCAATTCCTCCACAGATTTTGTTCAAAAT 1020  
DB 961 GGTTCCTCAATGTCCAGCTCCGGAAGGAATTTCAATTCCTCCACAGATTTTGTTCAAAAT 1020  
QY 1021 TAGTCTCTTTGGTTACAGCATGGATGATACCTGGGCGACCATTAATGCTCTACTTTTCCG 1080  
DB 1021 TAGTCTCTTTGGTTACAGCATGGATGATACCTGGGCGACCATTAATGCTCTACTTTTCCG 1080  
QY 1081 ATTAGTTCATCTCGAAGGAGAAATGTGGCATGCAATTCAGCATGCTTAGAGAACTGTA 1140  
DB 1081 ATTAGTTCATCTCGAAGGAGAAATGTGGCATGCAATTCAGCATGCTTAGAGAACTGTA 1140  
QY 1141 CTATATAAGAGAGTCTGTTATGATCCCGTGAGGTGGCTCACTGAGCAGTATGATGGTA 1200  
DB 1141 CTATATAAGAGAGTCTGTTATGATCCCGTGAGGTGGCTCACTGAGCAGTATGATGGTA 1200  
QY 1201 TCTCAAGGGTACAACTCCCAAAATCTCCGTCATCCTTTAGATGATGGTGGTGTGTA 1260  
DB 1201 TCTCAAGGGTACAACTCCCAAAATCTCCGTCATCCTTTAGATGATGGTGGTGTGTA 1260  
QY 1261 TGTAAAGAGGTCTAGTAACACCATGCAAAAGTTTATTTTGTGGTCCAGAGGTTAATGT 1320  
DB 1261 TGTAAAGAGGTCTAGTAACACCATGCAAAAGTTTATTTTGTGGTCCAGAGGTTAATGT 1320  
QY 1321 TTCCAATCGGTTCTCCGCAATTTATCTGAAGACATAGATAACTTCTTCTGTTCTTT 1380  
DB 1321 TTCCAATCGGTTCTCCGCAATTTATCTGAAGACATAGATAACTTCTTCTGTTCTTT 1380  
QY 1381 TGTGATCAGAGTGGGAGAACTGTATCTACAGACTTATACCAAGCAAGTACTG 1440  
DB 1381 TGTGATCAGAGTGGGAGAACTGTATCTACAGACTTATACCAAGCAAGTACTG 1440  
QY 1441 AAGTGTCTCAGGACAAACATCTATGAGAGATCTTATCAACTCTCGGGAAGCTTTGT 1500  
DB 1441 AAGTGTCTCAGGACAAACATCTATGAGAGATCTTATCAACTCTCGGGAAGCTTTGT 1500  
QY 1501 AATTGGTGTATATAAAATTTGAATTTCTTGCATTTTCATCGAGCCAGTTGCGGGATAATC 1560  
DB 1501 AATTGGTGTATATAAAATTTGAATTTCTTGCATTTTCATCGAGCCAGTTGCGGGATAATC 1560  
QY 1561 AGTGTGGATGTTGCATCAAGACCTGGCTTACTGCAATGAGATTAAGAGCTTGGATGG 1620  
DB 1561 AGTGTGGATGTTGCATCAAGACCTGGCTTACTGCAATGAGATTAAGAGCTTGGATGG 1620  
QY 1621 TGATTTTCGAGATCAAGATGTGCAAAATATGTCGCCAGACTTGGTCAATCTTTTGG 1680  
DB 1621 TGATTTTCGAGATCAAGATGTGCAAAATATGTCGCCAGACTTGGTCAATCTTTTGG 1680  
QY 1681 TTCTCCAGAGAGACTTTGAGTGTCTTAGCATGAGATGAAGTTATCCCGATGATAA 1740  
DB 1681 TTCTCCAGAGAGACTTTGAGTGTCTTAGCATGAGATGAAGTTATCCCGATGATAA 1740  
QY 1741 GGTTCATGGAACTGATGCTTTTCTGATGGAATTTGGTAAATATCTGGTACTTTGC 1800  
DB 1741 GGTTCATGGAACTGATGCTTTTCTGATGGAATTTGGTAAATATCTGGTACTTTGC 1800  
QY 1801 TCATAGAGTTCCTCAAAATGTCGCTTCAATATACCCCATCTCTCTCCAGATTCGTTA 1860  
DB 1801 TCATAGAGTTCCTCAAAATGTCGCTTCAATATACCCCATCTCTCTCCAGATTCGTTA 1860  
QY 1861 TSGTGGATATAAAGGTGTTGTGGGTGTTGATCCGGATTCATCAATGAAGTTGTTCTTGA 1920

1861 TGGTGATATAAGGTGTGTGGTGTGTGATCCGGATTCATCAATGAAGTTGTCTTGAG 1920  
1921 AAAGAGCATGTCGAAATATGAATCAGACACATCAATGAATGATGCTCCCTGGATCGACAA 1980  
1921 AAAGAGCATGTCGAAATATGAATCAGACACATCAATGAATGATGCTCCCTGGATCGACAA 1980  
1981 ATATCAGCCTTGTTATCTTAATCTGCAACTGATTACGCTCTGCTACACTGGAGTGAA 2040  
1981 ATATCAGCCTTGTTATCTTAATCTGCAACTGATTACGCTCTGCTACACTGGAGTGAA 2040  
2041 AGATGAAGTCTCCAGACAGACAAAGAGAGCTGTAGATCAGCTTGTATGCTATCTTGCA 2100  
2041 AGATGAAGTCTCCAGACAGACAAAGAGAGCTGTAGATCAGCTTGTATGCTATCTTGCA 2100  
2101 TGATTCCTTTGAAGCAGACAGGAGGCTTTGGAATGATGCTCTCCGAGAGAACACTAATAT 2160  
2101 TGATTCCTTTGAAGCAGACAGGAGGCTTTGGAATGATGCTCTCCGAGAGAACACTAATAT 2160  
2161 TCTCAAGGCAATGCTAACTGTGTTATAGCCCTGATGAGCCCTTCTTTCAATGAT 2220  
2161 TCTCAAGGCAATGCTAACTGTGTTATAGCCCTGATGAGCCCTTCTTTCAATGAT 2220  
2221 GTTGCAAAACCTCCGCGCATCCAAGTTGCTCGATTTGCGGACTAGATCAAGAATATTTAT 2280  
2221 GTTGCAAAACCTCCGCGCATCCAAGTTGCTCGATTTGCGGACTAGATCAAGAATATTTAT 2280  
2281 TCCAAATGGAAGAACATGATGGGATGTTTGGATGAATCCAGAACCTTGGAAATGCTCA 2340  
2281 TCCAAATGGAAGAACATGATGGGATGTTTGGATGAATCCAGAACCTTGGAAATGCTCA 2340  
2341 GGTGTTTGTTCAGTTTACTGTGCTGACATGGAGAGCTTTTCTGACGATTTTACATCAAT 2400  
2341 GGTGTTTGTTCAGTTTACTGTGCTGACATGGAGAGCTTTTCTGACGATTTTACATCAAT 2400  
2401 TAATAACAGCAGATCCACCAACAGTAATTTCAATCTCAAGGGAATGTTGTTGTGCAAA 2460  
2401 TAATAACAGCAGATCCACCAACAGTAATTTCAATCTCAAGGGAATGTTGTTGTGCAAA 2460  
2461 AAATCCATGCTGTATGCTGCTGATATCGTTTAAAGCGTGTAATGTTGAGCGCT 2520  
2461 AAATCCATGCTGTATGCTGCTGATATCGTTTAAAGCGTGTAATGTTGAGCGCT 2520  
2521 GCACCATAGTATGATGTTGTTATTCCTCCAGAAAGAAAGAGCTCATCCGAATGA 2580  
2521 GCACCATAGTATGATGTTGTTATTCCTCCAGAAAGAAAGAGCTCATCCGAATGA 2580  
2581 ATGTTCTGGAGTATGTTGATGGGATATCTACTTTGTTGTTGGATCAAGACATGAT 2640  
2581 ATGTTCTGGAGTATGTTGATGGGATATCTACTTTGTTGTTGGATCAAGACATGAT 2640  
2641 CCCGCCAAGGCAAGTCCAGCGGATGGAATATCCTCCAGCACCCAGCATACAGTTGGACCA 2700  
2641 CCCGCCAAGGCAAGTCCAGCGGATGGAATATCCTCCAGCACCCAGCATACAGTTGGACCA 2700  
2701 TGATGTCACAAATTGAGGAATTGAAGATGATCTCACCAACTATATTGTTGAATGACAGTTT 2760  
2701 TGATGTCACAAATTGAGGAATTGAAGATGATCTCACCAACTATATTGTTGAATGACAGTTT 2760  
2761 GGGATCATAGCAATCCCATGCTGCTGATTTGAGACAGAGAACCTGATATGCCCATGAG 2820  
2761 GGGATCATAGCAATCCCATGCTGCTGATTTGAGACAGAGAACCTGATATGCCCATGAG 2820  
2821 TGATTCATGCAAAATACTGCTGAGCTCTTTTCAATTCGAGTGGAGCTTTCCAAAGACTGG 2880  
2821 TGATTCATGCAAAATACTGCTGAGCTCTTTTCAATTCGAGTGGAGCTTTCCAAAGACTGG 2880  
2881 TGTTCGCGCTGAATACCATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2940  
2881 TGTTCGCGCTGAATACCATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2940  
2941 GCCGGACAAGACCACTATATCTCAGAAAGAGTTATTGAAAGCTTTTTCAGGAAAGTGAA 3000

2941 GCCGGACAAGACCACTATATCTCAGAAAGAGTTATTGAAAGCTTTTTCAGGAAAGTGAA 3000  
3001 GGACAAAGCACCTCCAGCTAGCTCTATCGCGACCTTCACAAAGAGATGTTGCAAGGAGATC 3060  
3001 GGACAAAGCACCTCCAGCTAGCTCTATCGCGACCTTCACAAAGAGATGTTGCAAGGAGATC 3060  
3061 ATATGATGCTGATATGGAAGTTGATGATTTGAAGATTTACATTGACGAAAGCTTTTGACAT 3120  
3061 ATATGATGCTGATATGGAAGTTGATGATTTGAAGATTTACATTGACGAAAGCTTTTGACAT 3120  
3121 CAAACTGATATGACAAAGCTGGTAAATTTAATGGACTACTATGGCATATAAAACAGA 3180  
3121 CAAACTGATATGACAAAGCTGGTAAATTTAATGGACTACTATGGCATATAAAACAGA 3180  
3181 GGCTGAAATACTTATGATGGTGGCATTTATGAAGCATCAAAAACCTTTTACCGCAGAAAAGA 3240  
3181 GGCTGAAATACTTATGATGGTGGCATTTATGAAGCATCAAAAACCTTTTACCGCAGAAAAGA 3240  
3241 TGCTGAGGCAATGATGTTGCTGTGAGGCGCTTGAGAGGAGGCAAGAGCCTGGTTCAA 3300  
3241 TGCTGAGGCAATGATGTTGCTGTGAGGCGCTTGAGAGGAGGCAAGAGCCTGGTTCAA 3300  
3301 GAGGCGCTAATGATATAGATGACATGTTACCAAGGCTTCGGCTTGGTACACGTTACATA 3360  
3301 GAGGCGCTAATGATATAGATGACATGTTACCAAGGCTTCGGCTTGGTACACGTTACATA 3360  
3361 TCATCTACATATTTGGGTTGCTTACAAATCAGGGGTTGAAAAGAGCTCATTTTATTAGCTT 3420  
3361 TCATCTACATATTTGGGTTGCTTACAAATCAGGGGTTGAAAAGAGCTCATTTTATTAGCTT 3420  
3421 TCCTCGTGTGTTTATGACCACTAATCCAGATTAAGAGGACAAAGACGTAACAGGCC 3480  
3421 TCCTCGTGTGTTTATGACCACTAATCCAGATTAAGAGGACAAAGACGTAACAGGCC 3480  
3481 AGTCTCAACTTGTCTATCTCTCAGGCGCTCACTAGTACAGATTAAGTGTGAAATGAGA 3540  
3481 AGTCTCAACTTGTCTATCTCTCAGGCGCTCACTAGTACAGATTAAGTGTGAAATGAGA 3540  
3541 TTCCAGTCCAGCGTTAAGCTGATATATATATGTAATAGGGTGTGATCATAAGAAAC 3600  
3541 TTCCAGTCCAGCGTTAAGCTGATATATATATGTAATAGGGTGTGATCATAAGAAAC 3600  
3601 TGTATGCAATGTTGACTACCTTTTGTCTTTTAAACTGCATGAAGCTGCAACATATATGC 3660  
3601 TGTATGCAATGTTGACTACCTTTTGTCTTTTAAACTGCATGAAGCTGCAACATATATGC 3660  
3661 AGTACTCTAAGAACAGATGTACAGCTAAGTACTAATATGATGATGATTTGATTTGATTC 3720  
3661 AGTACTCTAAGAACAGATGTACAGCTAAGTACTAATATGATGATTTGATTTGATTC 3720  
3721 TTTCTTCTAAA 3731  
3721 TTTCTTCTAAA 3731

RESULT 2

US-09-782-874-4/c  
; Sequence 4, Application US/09782874  
; Patent No. US20010023067A1  
GENERAL INFORMATION:

APPLICANT: Wassenecker, Michael  
Riedel, Leonhard  
Schiebel, Winfried  
Sanger, Heinz

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN  
RNA-DIRECTED RNA POLYMERASE (RdRp)

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York

```
;
; COUNTRY: USA
; ZIP: 10020
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/782,874
; FILING DATE: 08-Feb-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/811,583
; FILING DATE: 05-MAR-1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
;
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "primer"
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-782-874-4
;
; Query Match 0.7%; Score 25; DB 10; Length 25;
; Best Local Similarity 100.0%; Pred. No. 0.0044;
; Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1838 CCATCTGCTTCCAGATTCGTTATG 1862
;
; Db 25 CCATCTGCTTCCAGATTCGTTATG 1
;
; RESULT 3
;
; US-09-782-874-6/c
; Sequence 6, Application US/09782874
; Patent No. US20010023067A1
;
; GENERAL INFORMATION:
; APPLICANT: Wasseneegger, Michael
; Riedel, Leonhard
; Schiebel, Winfried
; Sanger, Heinz
;
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; POLYPEPTIDES HAVING THE ENZYMAIC ACTIVITY OF AN
; RNA-DIRECTED RNA POLYMERASE (RGRP)
;
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/782,874
; FILING DATE: 08-Feb-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/811,583
; FILING DATE: 05-MAR-1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
;
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
;
; US-09-782-874-5/c
; Sequence 5, Application US/09782874
; Patent No. US20010023067A1
;
; GENERAL INFORMATION:
; APPLICANT: Wasseneegger, Michael
; Riedel, Leonhard
; Schiebel, Winfried
; Sanger, Heinz
;
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; POLYPEPTIDES HAVING THE ENZYMAIC ACTIVITY OF AN
; RNA-DIRECTED RNA POLYMERASE (RGRP)
;
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/782,874
; FILING DATE: 08-Feb-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/811,583
; FILING DATE: 05-MAR-1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
;
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "primer"
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-782-874-6
;
; Query Match 0.7%; Score 25; DB 10; Length 25;
; Best Local Similarity 100.0%; Pred. No. 0.0044;
; Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 2659 GCCGATGGATATCTCTCCAGCACCC 2683
;
; Db 25 GCCGATGGATATCTCTCCAGCACCC 1
;
; RESULT 4
;
; US-09-782-874-5/c
; Sequence 5, Application US/09782874
; Patent No. US20010023067A1
;
; GENERAL INFORMATION:
; APPLICANT: Wasseneegger, Michael
; Riedel, Leonhard
; Schiebel, Winfried
; Sanger, Heinz
;
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; POLYPEPTIDES HAVING THE ENZYMAIC ACTIVITY OF AN
; RNA-DIRECTED RNA POLYMERASE (RGRP)
;
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/782,874
; FILING DATE: 08-Feb-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/811,583
; FILING DATE: 05-MAR-1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
;
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
```

Matches	20;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps
Qy	1916	TTGAGAAAGAGCATGTCGAA	1935					
Db	384	TTGAGAAAGAGCATGTCGAA	365					
RESULT 7								
US-09-998-598-772/c								
; Sequence 772, Application US/09998598								
; Patent No. US20020150922A1								
; GENERAL INFORMATION:								
; APPLICANT: Stolk, John A.								
; APPLICANT: Xu, Jiangchun								
; APPLICANT: Chenault, Ruth A.								
; APPLICANT: Meagher, Madelein Joy								
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND								
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER								
; FILE REFERENCE: 210121.561								
; CURRENT APPLICATION NUMBER: US/09/998,598								
; CURRENT FILING DATE: 2001-11-16								
; NUMBER OF SEQ ID NOS: 2606								
; SOFTWARE: Corixa Invention Disclosure Database								
; SEQ ID NO 772								
; TYPE: DNA								
; ORGANISM: Homo sapiens								
US-09-998-598-772								
Query Match								
Best Local Similarity 100.0%; Score 20; DB 10; Length 409;								
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0								
Qy	1916	TTGAGAAAGAGCATGTCGAA	1935					
Db	360	TTGAGAAAGAGCATGTCGAA	341					
RESULT 8								
US-09-998-598-2240/c								
; Sequence 2240, Application US/09998598								
; Patent No. US20020150922A1								
; GENERAL INFORMATION:								
; APPLICANT: Stolk, John A.								
; APPLICANT: Xu, Jiangchun								
; APPLICANT: Chenault, Ruth A.								
; APPLICANT: Meagher, Madelein Joy								
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND								
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER								
; FILE REFERENCE: 210121.561								
; CURRENT APPLICATION NUMBER: US/09/998,598								
; CURRENT FILING DATE: 2001-11-16								
; NUMBER OF SEQ ID NOS: 2606								
; SOFTWARE: Corixa Invention Disclosure Database								
; SEQ ID NO 2240								
; TYPE: DNA								
; ORGANISM: Homo sapiens								
US-09-998-598-2240								
Query Match								
Best Local Similarity 100.0%; Score 20; DB 10; Length 413;								
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0								
Qy	1916	TTGAGAAAGAGCATGTCGAA	1935					
Db	364	TTGAGAAAGAGCATGTCGAA	345					
RESULT 9								
US-09-840-787-54								
; Sequence 54, Application US/09840787								
; Patent No. US20020058264A1								

```

;
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; Hillman, Jennifer L.
; Bandman, Olga
; Shah, Purvi
; Au-Young, Janice
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
;
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/840,787
; FILING DATE: 23-Apr-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/518,865
; FILING DATE: <Unknown>
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX: <Unknown>
;
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1750 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; IMMEDIATE SOURCE:
; LIBRARY: HMCINOT01
; CLONE: 9476
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 54 :
US-09-840-787-54
;
; Query Match 0.5%; Score 20; DB 10; Length 1750;
; Best Local Similarity 100.0%; Pred. No. 3.2;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1916 TTGAGAAAGAGCATGTGCGAA 1935
; |||||
; DB 1355 TTGAGAAAGAGCATGTGCGAA 1374
;
; RESULT 10
; US-09-782-874-7
; Sequence 7, Application US/09782874
; Patent No. US20010023067A1
;
; GENERAL INFORMATION:
; APPLICANT: Wassenecker, Michael
; Riedel, Leonhard
; Schiebel, Winfried
; Sanger, Heinz
;
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; RNA-DIRECTED RNA POLYMERASE (RDRP)
;
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:

```

```

;
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/782,874
; FILING DATE: 08-Feb-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/811,583
; FILING DATE: 05-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
;
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "primer"
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-782-874-7
;
; Query Match 0.5%; Score 19; DB 10; Length 34;
; Best Local Similarity 100.0%; Pred. No. 7.3;
; Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 33 ACTCATCACTCCCTCAAG 51
; |||||
; DB 16 ACTCATCACTCCCTCAAG 34
;
; RESULT 11
; US-09-563-817-946/C
; Sequence 946, Application US/09563817
; Patent No. US20020095031A1
;
; GENERAL INFORMATION:
; APPLICANT: Nehls, Michael C.
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020095031A1 Human Polynucleotides and the
; FILE REFERENCE: LEX-0021-USA
; CURRENT APPLICATION NUMBER: US/09/563,817
; CURRENT FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: US 60/132,343
; PRIOR FILING DATE: 1999-05-04
; NUMBER OF SEQ ID NOS: 1008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 946
; LENGTH: 267
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(267)
; OTHER INFORMATION: n = A,T,C or G
;
; US-09-563-817-946
;
; Query Match 0.5%; Score 19; DB 10; Length 267;

```

Best Local Similarity 100.0%; Pred. No. 9;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 TCTTCTTTTGAAGGCTT 428  
|||||  
DB 22 TCTTCTTTTGAAGGCTT 4

RESULT 12  
US-09-864-761-11608/c  
; Sequence 11608, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecmica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 11608  
; LENGTH: 461  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC006204.1  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.96  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.73  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.96  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.79  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7  
US-09-864-761-11608

Query Match 0.5%; Score 19; DB 10; Length 461;  
Best Local Similarity 100.0%; Pred. No. 9.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 TCATATGAAATATATGGC 647  
|||||  
DB 56 TCATATGAAATATATGGC 38

RESULT 13  
US-09-764-877-2690  
; Sequence 2690, Application US/09764877  
; Patent No. US20020147140A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC005  
; CURRENT APPLICATION NUMBER: US/09/764,877  
; PRIOR FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 4031  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 2690  
; LENGTH: 4656  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-877-2690

Query Match 0.5%; Score 19; DB 10; Length 4656;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3567 TATATATGTAATAGGTG 3585  
|||||  
DB 2822 TATATATGTAATAGGTG 2840

RESULT 14  
US-09-782-874-9/c  
; Sequence 9, Application US/09782874  
; Patent No. US20010023067A1  
; GENERAL INFORMATION:  
; APPLICANT: Wassenecker, Michael  
; Riedel, Leonhard  
; Schiebel, Winfried  
; Sanger, Heinz  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN  
; RNA-DIRECTED RNA POLYMERASE (RDRP)  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & NEAVE  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/09/782,874  
; APPLICATION NUMBER: US/09/782,874  
; FILING DATE: 08-Feb-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/811,583  
; FILING DATE: 05-MAR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley, James F.  
; REGISTRATION NUMBER: 27,794

REFERENCE/DOCKET NUMBER: MFG-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-596-9000  
TELEFAX: 212-596-9090  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "primer"  
SEQUENCE DESCRIPTION: SEQ ID NO: 9:  
US-09-782-874-9

Query Match 0.5%; Score 18; DB 10; Length 35;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3615 GACTACCTTTGTCCTTA 3632  
|||||  
DB 35 GACTACCTTTGTCCTTA 18

RESULT 15  
US-09-923-876-1838/c  
; Sequence 1838, Application US/09923876  
; Patent No. US20020013958A1  
; GENERAL INFORMATION:  
; APPLICANT: Lalgudi, Raghunath V.  
; APPLICANT: Kamigaki, Laura Y. (Ito)  
; APPLICANT: Sherman, Bradley K.  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING  
; FILE REFERENCE: PL-0012-1 CON  
; CURRENT APPLICATION NUMBER: US/09/923,876  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: 09/298,329  
; PRIOR FILING DATE: 1999-04-21  
; PRIOR APPLICATION NUMBER: 60/085,331  
; PRIOR FILING DATE: 1998-05-05  
; NUMBER OF SEQ ID NOS: 6332  
; SOFTWARE: PERL Program  
; SEQ ID NO 1838  
; LENGTH: 256  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700159613H2  
US-09-923-876-1838

Query Match 0.5%; Score 18; DB 10; Length 256;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2120 GAGCGTTTGGATTGATG 2137  
|||||  
DB 80 GAGCGTTTGGATTGATG 63

RESULT 16  
US-09-960-352-1347/c  
; Sequence 1347, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960,352

CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 1347  
; LENGTH: 358  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 06-LIB34-055-Q1-E1-B5  
US-09-960-352-1347

Query Match 0.5%; Score 18; DB 10; Length 358;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 581 TTTTTCCTTTTACT 598  
|||||  
DB 263 TTTTTCCTTTTACT 246

RESULT 17  
US-09-867-701-4817  
; Sequence 4817, Application US/09867701  
; Patent No. US20020132237A1  
; GENERAL INFORMATION:  
; APPLICANT: Aglate, Paul A.  
; APPLICANT: Jones, Robert  
; APPLICANT: Harlocker, Susan L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER  
; FILE REFERENCE: 210121.497  
; CURRENT APPLICATION NUMBER: US/09/867,701  
; CURRENT FILING DATE: 2001-05-29  
; NUMBER OF SEQ ID NOS: 10912  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4817  
; LENGTH: 385  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-867-701-4817

Query Match 0.5%; Score 18; DB 10; Length 385;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 581 TTTTTCCTTTTACT 598  
|||||  
DB 114 TTTTTCCTTTTACT 131

RESULT 18  
US-09-864-761-11283/c  
; Sequence 11283, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/532,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666

Query Match	0.5%;	Score 18;	DB 10;	Length 474;
Best Local Similarity	100.0%;	Pred. No. 33;		
Matches 18;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 2280	TTCCAAATGGGAAGACAA 2297			
Db 32	TTCCAAATGGGAAGACAA 15			
RESULT 19				
US-09-783-590-8123/c				
Sequence 8123, Application US/09783590				
Patent No. US20020110850A1				
GENERAL INFORMATION:				
APPLICANT: Dillon, Patrick J.				
APPLICANT: Haseltine, William A.				
APPLICANT: Li, Haodong				
APPLICANT: Rosen, Craig A.				
APPLICANT: Ruben, Steven M.				
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2				
FILE REFERENCE: PO-16-2C1				
CURRENT APPLICATION NUMBER: US/09/783,590				
CURRENT FILING DATE: 2000-02-15				
PRIOR APPLICATION NUMBER: 08/420,856				
PRIOR FILING DATE: 1995-04-12				
PRIOR APPLICATION NUMBER: 08/346,731				
PRIOR FILING DATE: 1994-11-21				
NUMBER OF SEQ ID NOS: 12485				
SOFTWARE: Patentin Ver. 2.0				
SEQ ID NO 8123				
LENGTH: 500				
TYPE: DNA				
ORGANISM: Homo sapiens				
FEATURE:				
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.9				
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2				
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1				
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.86				
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8				
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4				
US-09-864-761-11283				
Query Match	0.5%;	Score 18;	DB 10;	Length 474;
Best Local Similarity	100.0%;	Pred. No. 33;		
Matches 18;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 2280	TTCCAAATGGGAAGACAA 2297			
Db 32	TTCCAAATGGGAAGACAA 15			
RESULT 19				
US-09-783-590-8123/c				
Sequence 8123, Application US/09783590				
Patent No. US20020110850A1				
GENERAL INFORMATION:				
APPLICANT: Dillon, Patrick J.				
APPLICANT: Haseltine, William A.				
APPLICANT: Li, Haodong				
APPLICANT: Rosen, Craig A.				
APPLICANT: Ruben, Steven M.				
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2				
FILE REFERENCE: PO-16-2C1				
CURRENT APPLICATION NUMBER: US/09/783,590				
CURRENT FILING DATE: 2000-02-15				
PRIOR APPLICATION NUMBER: 08/420,856				
PRIOR FILING DATE: 1995-04-12				
PRIOR APPLICATION NUMBER: 08/346,731				
PRIOR FILING DATE: 1994-11-21				
NUMBER OF SEQ ID NOS: 12485				
SOFTWARE: Patentin Ver. 2.0				
SEQ ID NO 8123				
LENGTH: 500				
TYPE: DNA				
ORGANISM: Homo sapiens				
FEATURE:				
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.9				
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2				
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1				
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.86				
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8				
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4				
US-09-864-761-11283				



```
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (433)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (455)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (461)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (467)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (468)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (473)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (476)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (477)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (479)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (480)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (484)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (486)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (487)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (495)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (497)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-783-590-8123
```

```
Query Match 0.5%; Score 18; DB 10; Length 500;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 576 AGAAATTTTCTTTT 593
    |||||
Db 70 AGAAATTTTCTTTT 53
```

## RESULT 20

```
; Sequence 176, Application US/09925299
; Patent No. US2002005627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 176
; LENGTH: 1016
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (895)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (928)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (970)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (992)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (1001)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-925-299-176
```

```
Query Match 0.5%; Score 18; DB 10; Length 1016;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 576 AGAAATTTTCTTTT 593
    |||||
Db 96 AGAAATTTTCTTTT 79
```

## RESULT 21

```
; Sequence 7347, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7347
; LENGTH: 1467
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1467)
; US-09-815-242-7347
```

```
Query Match          0.5%; Score 18; DB 10; Length 1467;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3628 CTTTAAACTGCATGAAG 3645
      |||||
Db 155 CTTTAAACTGCATGAAG 172

RESULT 22
US-09-887-576-629/c
; Sequence 629, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 629
; LENGTH: 1497
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-887-576-629

Query Match          0.5%; Score 18; DB 10; Length 1497;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1189 GTATGATGGGTATCTCAA 1206
      |||||
Db 645 GTATGATGGGTATCTCAA 628

RESULT 23
US-09-887-576-619/c
; Sequence 619, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 619
; LENGTH: 1497
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-887-576-619

Query Match          0.5%; Score 18; DB 10; Length 1944;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1189 GTATGATGGGTATCTCAA 1206
      |||||
Db 842 GTATGATGGGTATCTCAA 825

RESULT 24
US-09-841-132-328/c
; Sequence 328, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 328
; LENGTH: 2148
; TYPE: DNA
; ORGANISM: Chlamydia trachomatis
US-09-841-132-328

Query Match          0.5%; Score 18; DB 10; Length 2148;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3657 ATGCAGTACTCTAAGAAA 3674
      |||||
Db 1700 ATGCAGTACTCTAAGAAA 1683

RESULT 25
US-10-007-693-80/c
; Sequence 80, Application US/10007693
; Patent No. US20020146776A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT
; FILE REFERENCE: 210121.515C2
; CURRENT APPLICATION NUMBER: US/10/007,693
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 157
; SEQ ID NO 80
; LENGTH: 3048
; TYPE: DNA
; ORGANISM: Chlamydia trachomatis serovar D
US-10-007-693-80

Query Match          0.5%; Score 18; DB 12; Length 3048;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3657 ATGCAGTACTCTAAGAAA 3674
      |||||
Db 2603 ATGCAGTACTCTAAGAAA 2586
```

## RESULT 26

US-09-822-849A-162/c  
; Sequence 162, Application US/09822849A  
; Patent No. US20020045170A1  
; GENERAL INFORMATION:  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fechtel, Kim  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Howes, Steven H.  
; APPLICANT: Resnick, Richard J.  
; APPLICANT: Gulukota, Kamalakara  
; APPLICANT: Graham, James R.  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
; FILE REFERENCE: GIN 6403  
; CURRENT APPLICATION NUMBER: US/09/822,849A  
; CURRENT FILING DATE: 2001-09-04  
; PRIOR APPLICATION NUMBER: 60/195,582  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 598  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 162  
; LENGTH: 3760  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-822-849A-162

Query Match 0.5%; Score 18; DB 10; Length 3760;

Best Local Similarity 100.0%; Pred. No. 40;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 305 AACAGTCCAAAGGAGGA 322

Db 1490 AACAGTCCAAAGGAGGA 1473

## RESULT 27

US-10-052-586-453/c  
; Sequence 453, Application US/10052586  
; Patent No. US20020127584A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3430R1C1  
; CURRENT APPLICATION NUMBER: US/10/052,586  
; CURRENT FILING DATE: 2002-01-15  
; PRIOR APPLICATION NUMBER: 60/059263  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/059266  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/063120  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063121  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063486  
; PRIOR FILING DATE: 1997-10-21  
; PRIOR APPLICATION NUMBER: 60/063540  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063541

; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063544  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063564  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063734  
; PRIOR FILING DATE: 1997-10-29  
; PRIOR APPLICATION NUMBER: 60/063870  
; PRIOR FILING DATE: 1997-10-31  
; PRIOR APPLICATION NUMBER: 60/064103  
; PRIOR FILING DATE: 1997-10-31  
; PRIOR APPLICATION NUMBER: 60/065311  
; PRIOR FILING DATE: 1997-11-13  
; PRIOR APPLICATION NUMBER: 60/066120  
; PRIOR FILING DATE: 1997-11-21  
; PRIOR APPLICATION NUMBER: 60/066466  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/066772  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/069335  
; PRIOR FILING DATE: 1997-12-11  
; PRIOR APPLICATION NUMBER: 60/069425  
; PRIOR FILING DATE: 1997-12-12  
; PRIOR APPLICATION NUMBER: 60/069870  
; PRIOR FILING DATE: 1997-12-17  
; PRIOR APPLICATION NUMBER: 60/068017  
; PRIOR FILING DATE: 1997-12-18  
; PRIOR APPLICATION NUMBER: 60/077450  
; PRIOR FILING DATE: 1998-03-10  
; PRIOR APPLICATION NUMBER: 60/077632  
; PRIOR FILING DATE: 1998-03-11  
; PRIOR APPLICATION NUMBER: 60/077649  
; PRIOR FILING DATE: 1998-03-11  
; PRIOR APPLICATION NUMBER: 60/078886  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/078939  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/079664  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: 60/079786  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: 60/080107  
; PRIOR FILING DATE: 1998-03-31  
; PRIOR APPLICATION NUMBER: 60/080194  
; PRIOR FILING DATE: 1998-03-31  
; PRIOR APPLICATION NUMBER: 60/080327  
; PRIOR FILING DATE: 1998-04-01  
; PRIOR APPLICATION NUMBER: 60/080333  
; PRIOR FILING DATE: 1998-04-01  
; PRIOR APPLICATION NUMBER: 60/081049  
; PRIOR FILING DATE: 1998-04-08  
; PRIOR APPLICATION NUMBER: 60/081070  
; PRIOR FILING DATE: 1998-04-08  
; PRIOR APPLICATION NUMBER: 60/081195  
; PRIOR FILING DATE: 1998-04-09  
; PRIOR APPLICATION NUMBER: 60/081838  
; PRIOR FILING DATE: 1998-04-15  
; PRIOR APPLICATION NUMBER: 60/082568  
; PRIOR FILING DATE: 1998-04-21  
; PRIOR APPLICATION NUMBER: 60/082569  
; PRIOR FILING DATE: 1998-04-21  
; PRIOR APPLICATION NUMBER: 60/082704  
; PRIOR FILING DATE: 1998-04-22  
; PRIOR APPLICATION NUMBER: 60/082797  
; PRIOR FILING DATE: 1998-04-22  
; PRIOR APPLICATION NUMBER: 60/083322  
; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: 60/083495  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083496  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083499  
; PRIOR FILING DATE: 1998-04-29

;; PRIOR APPLICATION NUMBER: 60/083559  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/084366  
;; PRIOR FILING DATE: 1998-05-05  
;; PRIOR APPLICATION NUMBER: 60/084414  
;; PRIOR FILING DATE: 1998-05-06  
;; PRIOR APPLICATION NUMBER: 60/084639  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084640  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084643  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/085573  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085579  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085580  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085582  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085700  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/086023  
;; PRIOR FILING DATE: 1998-05-18  
;; PRIOR APPLICATION NUMBER: 60/086392  
;; PRIOR FILING DATE: 1998-05-22  
;; PRIOR APPLICATION NUMBER: 60/086486  
;; PRIOR FILING DATE: 1998-05-22  
;; PRIOR APPLICATION NUMBER: 60/087098  
;; PRIOR FILING DATE: 1998-05-28  
;; PRIOR APPLICATION NUMBER: 60/087208  
;; PRIOR FILING DATE: 1998-05-28  
;; PRIOR APPLICATION NUMBER: 60/087609  
;; PRIOR FILING DATE: 1998-06-02  
;; PRIOR APPLICATION NUMBER: 60/087759  
;; PRIOR FILING DATE: 1998-06-02  
;; PRIOR APPLICATION NUMBER: 60/087827  
;; PRIOR FILING DATE: 1998-06-03  
;; PRIOR APPLICATION NUMBER: 60/088025  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088028  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088029  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088033  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088167  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088202  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088212  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088217  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088326  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088655  
;; PRIOR FILING DATE: 1998-06-09  
;; PRIOR APPLICATION NUMBER: 60/088722  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088738  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088740  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088811  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088824  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088825  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088826  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088861

;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/088863  
;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/088876  
;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/089090  
;; PRIOR FILING DATE: 1998-06-12  
;; PRIOR APPLICATION NUMBER: 60/089105  
;; PRIOR FILING DATE: 1998-06-12  
;; PRIOR APPLICATION NUMBER: 60/089512  
;; PRIOR FILING DATE: 1998-06-16  
;; PRIOR APPLICATION NUMBER: 60/089514  
;; PRIOR FILING DATE: 1998-06-16  
;; PRIOR APPLICATION NUMBER: 60/089538  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089598  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089653  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089908

Query Match 0.5%; Score 18; DB 12; Length 3781;  
Best Local Similarity 100.0%; Pred No. 40;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3267 GGGCCTTGAGGAAGAGG 3284  
|||||  
Db 1159 GGGCCTTGAGGAAGAGG 1142

## RESULT 28

US-09-879-445-1  
; Sequence 1, Application US/09879445  
; Patent No. US20020061509A1

## GENERAL INFORMATION:

APPLICANT: Mundy, Gregory R.

INVENTOR: Gallwitz, Wolf

TITLE OF INVENTION: SCREENING ASSAY FOR THE IDENTIFICATION  
OF AGENTS WHICH INHIBIT CANCER  
METASTASIS TO BONE

## NUMBER OF SEQUENCES: 1

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION NUMBER: US/09/879,445

FILING DATE: 11-Jun-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/915,868

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Highlander, Steven L.

REGISTRATION NUMBER: 37,642

REFERENCE/DOCKET NUMBER: OST:002PZ1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 512/418-3000

TELEFAX: 512/474-7577

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4348 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-879-445-1

Query Match 0.5%; Score 18; DB 10; Length 4348;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 576 AGAAATTTTCTTTT 593  
|||||  
DB 4180 AGAAATTTTCTTTT 4197

RESULT 29

US-09-764-877-3657  
; Sequence 3657, Application US/09764877

; Patent No. US20020147140A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC005

; CURRENT APPLICATION NUMBER: US/09/764,877

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - refer to PALM or file wrapper

; NUMBER OF SEQ ID NOS: 4031

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3657

; LENGTH: 32192

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-764-877-3657

Query Match 0.5%; Score 18; DB 10; Length 32192;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 TTAAGTCATCTTAGAGA 266  
|||||  
DB 28605 TTAAGTCATCTTAGAGA 28622

RESULT 30

US-09-927-091-4

; Sequence 4, Application US/09927091

; Patent No. US20020119541A1

; GENERAL INFORMATION:

; APPLICANT: KILLARY, ANN

; APPLICANT: LOTT, STEVE

; APPLICANT: CHANDLER, DAWN

; TITLE OF INVENTION: THE TUMOR SUPPRESSOR CAR-1

; FILE REFERENCE: UTSC:65105

; CURRENT APPLICATION NUMBER: US/09/927,091

; CURRENT FILING DATE: 2001-08-09

; PRIOR APPLICATION NUMBER: 60/227,560

; PRIOR FILING DATE: 2000-08-23

; PRIOR APPLICATION NUMBER: 60/225,033

; PRIOR FILING DATE: 2000-08-10

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4

; LENGTH: 49744

; TYPE: DNA

; ORGANISM: Human

US-09-927-091-4

Query Match 0.5%; Score 18; DB 10; Length 49744;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 581 TTTTTCCTTTTATCT 598  
|||||  
DB 1050 TTTTTCCTTTTATCT 1067

RESULT 31

US-09-835-232-7/c

; Sequence 7, Application US/09835232

; Patent No. US20020098489A1

; GENERAL INFORMATION:

; APPLICANT: Leder, Philip

; APPLICANT: Leder, Benjamin

; TITLE OF INVENTION: FORMIN-2 NUCLEIC ACIDS AND POLYPEPTIDES

; FILE REFERENCE: 00383/052002

; CURRENT APPLICATION NUMBER: US/09/835,232

; CURRENT FILING DATE: 2001-04-12

; PRIOR APPLICATION NUMBER: US 60/196,811

; PRIOR FILING DATE: 2000-04-13

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7

; LENGTH: 170834

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(170834)

; OTHER INFORMATION: n= A,T,C, or G

US-09-835-232-7

Query Match 0.5%; Score 18; DB 10; Length 170834;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 523 GTTTCAGTGTGGGAAG 540  
|||||

DB 36324 GTTTCAGTGTGGGAAG 36307

RESULT 32

US-09-923-876-3754

; Sequence 3754, Application US/09923876

; Patent No. US20020013958A1

; GENERAL INFORMATION:

; APPLICANT: Lalgudi, Raghunath V.

; APPLICANT: Kamigaki, Laura Y. (Ito)

; APPLICANT: Sherman, Bradley K.

; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING

; FILE REFERENCE: PL-0012-1 CON

; CURRENT APPLICATION NUMBER: US/09/923,876

; CURRENT FILING DATE: 2001-08-06

; PRIOR APPLICATION NUMBER: 09/298,329

; PRIOR FILING DATE: 1999-04-21

; PRIOR APPLICATION NUMBER: 60/085,331

; PRIOR FILING DATE: 1998-05-05

; NUMBER OF SEQ ID NOS: 6332

; SOFTWARE: PERL Program

; SEQ ID NO 3754

; LENGTH: 150

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; NAME/KEY: misc\_feature

; OTHER INFORMATION: Incyte ID No. US20020013958A1 700453751H1

; NAME/KEY: unsure

; LOCATION: 115, 118, 121, 129, 138, 147

; OTHER INFORMATION: a, t, c, g, or other

US-09-923-876-3754

Query Match 0.5%; Score 17; DB 10; Length 150;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2464 TCCATGCTTGATCTG 2480  
|||||  
DB 73 TCCATGCTTGATCTG 89

```
RESULT 33
US-09-878-574-9084
; Sequence 9084, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 9084
; LENGTH: 213
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701102066H1
US-09-878-574-9084

Query Match      0.5%  Score 17; DB 10; Length 213;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2745 TTGTGAATGACAGTTTG 2761
|||||
DB 125 TTGTGAATGACAGTTTG 141

RESULT 34
US-09-878-574-7593/c
; Sequence 7593, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 7593
; LENGTH: 258
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701100084H2
US-09-878-574-7593

Query Match      0.5%  Score 17; DB 10; Length 258;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2429 TTCATCTGAAGGGAAA 2445
|||||
DB 66 TTCATCTGAAGGGAAA 50

RESULT 35
US-09-969-373-41/c
; Sequence 41, Application US/09969373
; Patent No. US20020133852A1
; GENERAL INFORMATION:
; APPLICANT: Effertz, Roger J.
; APPLICANT: Hauge, Brian M.
; TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
```

```
; FILE REFERENCE: 38-10(52679)A
; CURRENT APPLICATION NUMBER: US/09/969,373
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US 09/754,853
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 09/760,427
; PRIOR FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: US 09/855,768
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 41
; LENGTH: 269
; TYPE: DNA
; ORGANISM: Glycine max
US-09-969-373-41

Query Match      0.5%  Score 17; DB 10; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 582 TTTTTCCTTTTATCT 598
|||||
DB 148 TTTTTCCTTTTATCT 132

RESULT 36
US-09-294-093B-1096
; Sequence 1096, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Lalquidi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 1096
; LENGTH: 274
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700343770H1
US-09-294-093B-1096

Query Match      0.5%  Score 17; DB 10; Length 274;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2567 CCTCATCCGAATGATG 2583
|||||
DB 128 CCTCATCCGAATGATG 144

RESULT 37
US-09-864-761-31520/c
; Sequence 31520, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
```

```
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31520
; LENGTH: 308
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006323.3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: NT HIT: U52965.1, EVALUE 1.00e-108
; OTHER INFORMATION: EST HUMAN HIT: BF026728.1, EVALUE 1.00e-108
; OTHER INFORMATION: SWISSPROT HIT: Q15910, EVALUE 6.00e-13
US-09-864-761-31520

Query Match 0.58; Score 17; DB 10; Length 308;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 3712 AGTTTCATCTTTCTCT 3728
Db 228 AGTTTCATCTTTCTCT 212

RESULT 38
US-09-880-107-706/c
; Sequence 706, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darcie T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherif, Owe

US-09-917-800A-623/c
; Sequence 623, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 623
; LENGTH: 353
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 AI045253
; NAME/KEY: misc.feature
; LOCATION: (1)..(353)
; OTHER INFORMATION: n = a o r c o r g o r t
US-09-917-800A-623
```

Query Match 0.5%; Score 17; DB 10; Length 353;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 AAATTTTCTTTTCTTTT 594  
|||||  
DB 64 AAATTTTCTTTTCTTTT 48

RESULT 40  
US-09-864-761-13956/c  
; Sequence 13956, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 13956  
; LENGTH: 368  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AFL30350.1  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3  
US-09-864-761-13956

Query Match 0.5%; Score 17; DB 10; Length 368;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 TTTATTTTGCATACTT 132  
|||||  
DB 236 TTTATTTTGCATACTT 220

RESULT 41  
US-09-974-300-7289/c  
; Sequence 7289, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; APPLICANT: Clausen, Ib Groth  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; TITLE OF INVENTION: Expression  
; FILE REFERENCE: 10095.500-US  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7289  
; LENGTH: 387  
; TYPE: DNA  
; ORGANISM: Bacillus clausii  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(387)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-974-300-7289

Query Match 0.5%; Score 17; DB 10; Length 387;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 AAATTTTCTTTTCTTTT 594  
|||||  
DB 268 AAATTTTCTTTTCTTTT 252

RESULT 42  
US-09-864-761-30520/c  
; Sequence 30520, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667



;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 30520  
;; LENGTH: 391  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AF130350.1  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3  
;; OTHER INFORMATION: NT HIT: AL163248.2, EVALU0 0.00e+00  
;; OTHER INFORMATION: EST\_HUMAN HIT: AA001786.1, EVALU0 3.00e-84  
;; OTHER INFORMATION: SWISSPROT HIT: P70213, EVALU0 6.00e-21  
US-09-864-761-30520

Query Match 0.5%; Score 17; DB 10; Length 391;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 TTTATTTTGCATAACTT 132  
|||||  
DB 81 TTTATTTTGCATAACTT 65

RESULT 43  
US-09-563-817-878/c  
; Sequence 878, Application US/09563817  
; Patent No. US20020095031A1  
; GENERAL INFORMATION:  
; APPLICANT: Nehls, Michael C.  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; TITLE OF INVENTION: NO. US20020095031A1 Human Polynucleotides and the  
; TITLE OF INVENTION: Polypeptides Encoded Thereby  
; FILE REFERENCE: LEX-0021-USA  
; CURRENT APPLICATION NUMBER: US/09/563,817  
; PRIOR FILING DATE: 2001-06-11  
; PRIOR APPLICATION NUMBER: US 60/132,343  
; PRIOR FILING DATE: 1999-05-04  
; NUMBER OF SEQ ID NOS: 1008  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 878  
; LENGTH: 391  
; TYPE: DNA  
; ORGANISM: homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)-(391)  
; OTHER INFORMATION: n - A,T,C or G

US-09-563-817-878

Query Match 0.5%; Score 17; DB 10; Length 391;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 754 CTTTAAGGAACCTCGT 770  
|||||  
DB 118 CTTTAAGGAACCTCGT 102

RESULT 44

US-09-983-965-401  
; Sequence 401, Application US/09983965  
; Patent No. US20020137160A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION  
; FILE REFERENCE: 37-21(10297)C  
; CURRENT APPLICATION NUMBER: US/09/983,965  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: US 09/465,231  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: US 60/113,678  
; PRIOR FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 5912  
; SEQ ID NO 401  
; LENGTH: 406  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 48-BOVMS1-008-Q1-E1-D12  
US-09-983-965-401

Query Match 0.5%; Score 17; DB 10; Length 406;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2494 TTTAAAGGCTGTAATG 2510  
|||||  
DB 41 TTTAAAGGCTGTAATG 57

RESULT 45

US-09-867-550-1095  
; Sequence 1095, Application US/09867550  
; Patent No. US20020082206A1  
; GENERAL INFORMATION:  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Mehraban, Fuad,  
; APPLICANT: Conley, Pamela  
; APPLICANT: Law, Debbie  
; APPLICANT: Topper, James  
; TITLE OF INVENTION: NO. US20020082206A1 Polynucleotides from Atherogenic Cells a  
; TITLE OF INVENTION: Thereby  
; FILE REFERENCE: 21402-013 (Cura-313)  
; CURRENT APPLICATION NUMBER: US/09/867,550  
; CURRENT FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: USN 60/208,427  
; PRIOR FILING DATE: 2000-05-30  
; NUMBER OF SEQ ID NOS: 2125  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1095  
; LENGTH: 411  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (336)..(359)

; OTHER INFORMATION: Wherein any n is one of a or t or c or g  
US-09-867-550-1095

Query Match 0.5%; Score 17; DB 10; Length 411;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1508 GATAAAATTTGAATT 1524  
|||||

DB 201 GATAAAATTTGAATT 217

## RESULT 46

US-09-960-352-10698  
; Sequence 10698, Application US/09960352  
; Patent No. US20020137139A1

## GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing

; APPLICANT: Byatt, John C.

; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960.352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 10698

; LENGTH: 419

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 46-LIB3057-002-Q1-K1-D6

US-09-960-352-10698

## Query Match

Best Local Similarity 0.5%; Score 17; DB 10; Length 419;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 580 ATTTTTCCTTTTAT 596  
|||||

DB 353 ATTTTTCCTTTTAT 369

## RESULT 47

US-09-880-107-2705

; Sequence 2705, Application US/09880107

; Patent No. US20020142981A1

## GENERAL INFORMATION:

; APPLICANT: Horne, Darci T.

; APPLICANT: Vockley, Joseph G.

; APPLICANT: Scherf, Uwe

; APPLICANT: Gene Logic, Inc.

; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer

; FILE REFERENCE: 44921-5028-WO

; CURRENT APPLICATION NUMBER: US/09/880.107

; CURRENT FILING DATE: 2001-06-14

; PRIOR APPLICATION NUMBER: US 60/211,379

; PRIOR FILING DATE: 2000-06-14

; PRIOR FILING DATE: 2000-06-14

; PRIOR FILING DATE: 2000-10-02

; NUMBER OF SEQ ID NOS: 3950

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2705

; LENGTH: 421

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Genbank Accession No. US20020142981A1 N81036

US-09-880-107-2705

## Query Match

Best Local Similarity 0.5%; Score 17; DB 10; Length 421;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1954 AAAGTTAGATGTCCTTG 1970  
|||||

DB 388 AAAGTTAGATGTCCTTG 404

## RESULT 48

US-09-864-761-14986/c

; Sequence 14986, Application US/09864761

; Patent No. US20020048763A1

## GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Weisheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO

; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

; FILE REFERENCE: Aeomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864.761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117

; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 14986

; LENGTH: 430

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: MAP TO AC006323.3

; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.9

; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.95

; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1

; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.7

; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5

; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2

; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2

US-09-864-761-14986

```

Query Match      0.5%; Score 17; DB 10; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3712 AGTTTCATCTTCTCT 3728
DB 405 AGTTTCATCTTCTCT 389
|||||
|||||

RESULT 49
US-09-864-761-2826
Sequence 2826, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 4917
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 2826
LENGTH: 436
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AJ229041.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9

```

```

; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6
US-09-864-761-2826

Query Match          0.5%; Score 17; DB 10; Length 436;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels

QY 579 AATTTTTCCTTTT 595
      |||
DB 126 AATTTTTCCTTTT 142

RESULT 50
US-09-983-965-2934/C
; Sequence 2934, Application US/0983965 .
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 2934
; LENGTH: 436
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 19-LI33058-019-Q1-K1-E11
US-09-983-965-2934

Query Match          0.5%; Score 17; DB 10; Length 436;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels

QY 578 AATTTTTCCTTTT 594
      |||
DB 157 AATTTTTCCTTTT 141

Search completed: November 6, 2002, 03:42:03
Job time : 850 secs

```

Search completed: November 6, 2002, 03:42:03  
Job time : 850 secs